Measuring economic uncertainty using news-media textual data

Peter, Eckley

University of Oxford, Bank of England

30 January 2015

Measuring economic uncertainty using news-media textual data

An improved methodology and a detailed comparison to stock returns volatility

Draft – comments welcome

May 2015

Peter Eckley, Department of Economics, University of Oxford, and Bank of England

(Supervised by Professor Christopher Bowdler, Department of Economics, University of Oxford)

Correspondence address: peter.eckley@bankofengland.co.uk

Abstract: We develop a news-media textual measure of aggregate economic uncertainty, defined as the fraction of Financial Times articles that contain uncertainty-related keyphrases, at frequencies from daily to annual, from January 1982 to April 2014.

We improve on existing similar measures in several ways. First, we reveal extensive and irregular duplication of articles in the news database most widely used in the literature, and provide a simple but effective de-duplication algorithm. Second, we boost the uncertainty 'signal strength' by 14% through the simple addition of the word "uncertainties" to the conventional keyword list of "uncertain" and "uncertainty", and show that adding further uncertainty-related keyphrases would likely constitute only a second-order adjustment. Third, we demonstrate the importance of normalising article counts by total news volume and provide the first textual uncertainty measure to do so for the UK.

We empirically establish the plausibility of our measure as an uncertainty proxy through a detailed narrative analysis and a detailed comparative analysis with another popular uncertainty proxy, stock returns volatility. We show the relationship between these proxies is strong and significant on average, but breaks down periodically. We offer plausible explanations for this behaviour. We also establish the absence of Granger causation between the measures, even down to daily (publication) frequency.

Contents

1 Introduction

Uncertainty is widely believed to affect a wide range of economic and financial decisions and outcomes, especially those involving a long horizon, such as company capital investment and hiring, durables consumption, precautionary savings, credit creation and long-term financial investment by households.

Unfortunately, uncertainty is not directly observed. A variety of uncertainty proxies have therefore been proposed for use in empirical work, including volatility (realised or options-implied) of financial returns or of aggregate economic indicators, cross-sectional dispersion of (disagreement between) forecasts, measures of scale of distributional/density forecasts, revision volatility of forecasts, size of forecast errors, and frequency of particular internet searches (see Nicholas Bloom (2014) for an overview). However there has been relatively little analysis of the properties of and relationship between these proxies.

A new class of uncertainty proxies has recently been proposed, based on automated analysis of news-media textual data (Alexopoulos & Cohen, 2009; Baker, Bloom, & Davis, 2013). Whether expressions of uncertainty in the news-media cause or only reflect the uncertainty of economic decision-makers, they may be an efficient way of capturing latent uncertainty. However, the attributes of these measures, and important aspects of their construction are as yet little explored, so that the growing literature that deploys these uncertainty proxies lacks a robust empirical foundation.

This paper begins to address some of these gaps in the literature, with contributions in three areas.

First, we provide a general framework for measuring latent uncertainty about a specified subject using news-media textual data. The basic idea is to count articles containing uncertainty-related keyphrases in a given period, and normalise this by the total number of articles in the period, to give the unitless fraction of articles that express uncertainty. The framework links observed article counts to latent uncertainty.

Second, within the above framework we provide an empirical implementation of a news-media textual measure of aggregate uncertainty and apply it to almost two million Financial Times articles, 1982-2013. We improve on the extant literature in two main ways. First, we show that duplication of articles is common in the raw data that is widely used. We identify patterns in the duplication and develop a de-duplication algorithm that exploits these patterns to remove duplicates. Second, we document both high frequency volatility and low frequency trends in news volume, driven by factors other than uncertainty. This highlights the importance of normalising article counts or frequencies by overall news volume in order to mitigate spurious movement that would otherwise appear as noise in the uncertainty measure. We present the first measure to do so on UK data.

Third, we conduct the first detailed comparative study of news-media uncertainty and stock returns volatility. This includes graphical, narrative, and correlational analyses and Granger causation testing at up to daily frequency. Our results provide a closer look at the strong relationship between the two measures that has been noted briefly in previous literature, showing the strength of relationship at frequencies up to daily, and in connection to key narrative events. However, our analysis also reveals that the long-term average correlation obscures switching behaviour between periods of high vs. low correlation. We advance hypotheses for why this might be.

The rest of this paper is structured as follows. Section 2 highlights some *a priori* merits of a news-media textual measure as a proxy for uncertainty, relative to stock volatility. Section 3 briefly reviews the related literature. Section 4 lays out our measurement framework and Section 5 provides our empirical implementation. Section 6 presents our comparative analysis. Section 7 draws conclusions and Section 8 suggests directions for future research.

2 Relative merits of a news-media textual uncertainty measure

Different uncertainty measures are likely to be complementary as much as competitors. However, it is worth noting some basic *a priori* advantages that news-media textual uncertainty measures possess over measures derived from financial asset prices, such as the standard deviation of realised stock returns which we compare with our news-media measure in this paper.

First, textual data provides *direct* unmediated expressions of uncertainty – most obviously in the use of the word "uncertainty". By contrast many assumptions are needed to extract an estimate of perceived uncertainty from financial asset price volatilities (or implied volatility) since these only *indirectly* affected by uncertainty – being the result of many individual decisions which may be influenced by other factors, such as time-varying risk aversion and liquidity preferences, institutional factors such as regulations and market structure, and norms/conventional heuristics. Furthermore, the semantics of "uncertainty" have changed little over recent history, whereas the mediating effect of non-uncertainty factors on the link between volatility and uncertainty may well have changed in the face of changes to market structure, technology, and financial regulation.

Second, textual expressions of uncertainty are often accompanied by richer information on the nature of uncertainty and its anticipated effects. This could be used to decompose aggregate uncertainty into components reflecting uncertainty about particular subjects, which could in turn be recombined into different indices relevant to particular decision contexts. Of course automated extraction of such structured information from natural language statements is challenging but in principle achievable, and the subject of much computational linguistics research. The scope for similar decompositions of financial volatility is typically more limited – being dependent on the existence of suitable financial instruments or contingent on modelling assumptions that are difficult to test[1].

Third, textual expressions of uncertainty are intrinsically point-in-time (the moment of publication). By contrast realised volatility is latent and must be inferred from sampling returns over extended periods. This issue may be substantially mitigated by intra-day sampling, or avoided entirely by estimating implied volatility from pricing of stock options and variance swaps. However, the necessary is available for fewer stocks than are covered by the news-media, and data availability is usually poor or non-existent prior to the 1990s or 2000s.

Among the many textual corpora that could be examined, news-media text is of particular interest for at least two reasons. First, its wide audience makes it a potential nexus in social processes of opinion and sentiment formation/transmission. Second, its consistent publication schedule, audience and content focus, and format makes it easier to construct measures that are comparable

[1] Admittedly, this is a prospective relative merit that is not fully realised in the present work. In Chapter 2 we decompose aggregate news-media uncertainty to a company-level but a similar decomposition can also be achieved using individual company stock returns. Nevertheless, news-media textual data should in principle admit greater granularity and flexibility of decomposition than would typically be possible using financial data.

from one period to the next than is the case with less structured corpora such as Twitter feeds, blogs or text from web crawlers.

3 Literature review

The literature on news-media measures of uncertainty is still nascent. The seminal work is Alexopoulos & Cohen (2009), which measured the frequency of articles in the New York Times containing the keywords “uncertainty” (or “uncertain”) and “economic” (or “economy”) for 1929-2008 (thus missing most of the recent crisis period). The authors contrast the effect of news-media uncertainty with that of stock volatility in separate low dimensional VARs and together in a trivariate VAR with various US output variables. This implies an *indirect* comparison of the measures, but the authors do not go beyond a cursory verbal analysis in comparing the two uncertainty proxies *directly*.

The next most closely related work is Baker et al. (2013). This focuses on US economic *policy* uncertainty and so counts only articles that contain particular policy-related keyphrases in addition to “uncertain” or “uncertainty”. The authors’ comparison of this measure to the US implied stock returns volatility index known as the VIX is limited to a graphical comparison of the movements of this measure around a small subset of large stock market index jumps and reporting the full sample Pearson’s correlation of 0.578[2].

A couple of recent papers have used news-media uncertainty measures as one component in composite uncertainty indices, deployed in macroeconomic VARs for the UK. However, these papers contain very limited analysis of the composite measure itself or of the relationship between the news-media component and the other components (Dendy, Mumtaz, & Silver, 2013; Haddow, Hare, Hooley, & Shakir, 2013).

To the best of our knowledge there has been no detailed study of key aspects of the construction methodology including de-duplication and normalisation by news volume. Nor has there been any detailed comparison of a news-media uncertainty measure with common alternative uncertainty proxies, such as stock volatility.

A larger parallel literature seeks to extract measures of general tone or sentiment (rather than uncertainty specifically) from news-media textual data using larger dictionaries of keywords. The focus is typically on predicting the level of stock returns, though a couple of papers have examined the link between sentiment measures (not uncertainty) and volatility (e.g. Kothari, Li, & Short, 2009; Tetlock, 2007). The only sentiment analysis work specifically on uncertainty that we know of is Loughran & McDonald (2011), which builds a dictionary of 285 words “denoting uncertainty, with emphasis on the general notion of imprecision rather than exclusively focussing on risk”. The authors examine the relationship between frequency of these words in the text of US company 10-K filings and changes in company stock returns, volume, and returns volatility around the filing publication date. They find a significant positive correlation between post-publication return volatility and occurrence of uncertainty words[3].

[2] They also report a correlation of 0.733 using a variant of their index focussed on articles containing terms that they *a priori* relate to the equity markets, rather than to economic policy *per se*.

[3] Our tentative understanding of their results (they do not provide an extended discussion or sufficient detail to construct a definitive interpretation for oneself) is that a one standard deviation increase in the frequency

Finally, our comparative analysis also connects to the literature on the correlates and determinants of stock returns volatility. In particular, if news-media uncertainty is interpreted as reflecting fundamentals then its correlation with volatility would support the hypothesis that the volatility is partly determined by fundamentals. Methodologically, parts of our quantitative analysis bear similarities to Campbell, Lettau, Malkiel, & Xu (2001) which considers the relationship between disaggregated components of aggregate stock returns volatility.

4 Measurement framework

The first step in measuring uncertainty is to define what we mean by it. This is the subject of Section 4.1. Section 4.2 briefly discusses the implications of using a scalar representation of uncertainty. Section 4.3 explains our framework for inferring latent uncertainty from textual data. Estimation of stock returns volatility, to be compared with news-media uncertainty, is deferred to Section 6.1.2.

While the empirical analysis in the present Chapter is focused on aggregate uncertainty, the framework we introduce in this Section will be cast more generally so that we can re-use it in the context of firm-level uncertainty measures in Chapter 2.

4.1 Defining uncertainty

One common approach to defining an economic variable to be measured – here uncertainty – is to do so *indirectly* by identifying it with a particular shock in a structural economic model, and thus delegating the definition and interpretation of that variable to the definition and interpretation of the structural shock.

However, this approach is not appropriate here. Most widely-used uncertainty proxies are often deployed in reduced-form models – whether macro VAR or micro-econometric models – where the correspondence to particular shocks is indeterminate and drawn loosely by narrative assertion. We envisage our news-media uncertainty measure being used in a similar way. Furthermore, we do not have a single model or decision context in mind in this Chapter.

An alternative approach is to define *directly* what we mean by uncertainty. We distinguish two dimensions of this definition[4].

The first is what the uncertainty is *about*, i.e. the subject of the uncertainty. Uncertainty can attach to many aspects of the economy, so in a sense there are as many relevant uncertainties or components of uncertainty as there are decision-relevant variables in economically-relevant decision contexts. In this sense, we should not speak of a single 'uncertainty' or subject or component of uncertainty, but rather should make clear the subject(s), people's uncertainty about which we are measuring. For example, the Financial Times covers a wide range of subjects whose uncertainty we will thus be measuring by considering all articles together, but in Chapter 2 we limit the subject focus to a given company by considering only the articles that discuss that company.

---

of uncertainty-related keywords in a 10-K is associated with a 10 percentage point increase in post-publication return volatility, which would be quite substantial compared to a typical annualised volatility of 10 to 20 percentage points.

[4] One could add a third dimension – *whose* uncertainty – in recognition that uncertainty, in the sense most relevant to human decision-making, is a perception of human beings, and human perceptions are often heterogeneous.

The second and more fundamental dimension in directly defining uncertainty is the essential nature of uncertainty *per se*. The reduction of uncertainty to actuarial or stochastic risk, as a modelling device, is so pervasive in modern economics that it is easily forgotten that other, ontologically different, forms or aspects of uncertainty exist, including ambiguity or Knightian (1921) uncertainty. On this front we suggest that the definition of uncertainty that is relevant for economic decision-making is inextricably bound up with the cognitive models of the economic actors making those decisions. Uncertainty is thus perhaps best understood as a cognitive state.

Cognitive state in economic decision contexts is not yet well observed, and we lack good models of the corresponding cognition. However, we would assert that one of the most proximate expressions of cognitive state is in tellingly-named natural language (more so than in the formal frameworks of stochastic risk or ambiguity). Therefore natural language expressions of uncertainty should certainly be considered objects of interest, as part of a progressive research program towards a richer understanding of economic cognition and decision-making under uncertainty.

4.2 Representing uncertainty

We represent latent uncertainty about subject i by the real-valued scalar, as is common practice in many parts of the economic literature. More generally, one might think of a vector of uncertainty components. Our scalar can be thought of either as an element in such a vector, or a weighted index of such elements. We normalise the scalar to the unit interval, and label it $U_{it}^* \in [0,1]$ where t indexes time periods. This provides a natural mapping to the empirical measure developed in this Chapter which is expressed as a fraction (of articles classified as expressing uncertainty) which lies between 0 and 1 inclusive.

In the present paper we consider aggregate uncertainty at a range of frequencies, so i includes all subjects about which uncertainty is expressed in the text of the Financial Times, and the periodicity of t ranges from one day to one calendar year. In most of this Chapter we omit the subscripts for the sake of brevity and denote latent aggregate uncertainty by U^*.

If we accord U^* only ordinal interpretation, such that U^* is monotonically increasing in latent uncertainty but the specific monotonic transformation function is unknown then the range restriction to the unit interval is merely a normalisation without economic content since a potential infinity of ranks can be expressed within any non-degenerate interval of the real line.

If interpretation is extended to include cardinality then the lower bound of zero has a natural interpretation as the complete absence of uncertainty (or, conversely, complete certainty) even if this is only a hypothetical limit that is unachievable in practice. It is less clear whether uncertainty can be bounded above in a cardinal sense, but the bounding above by 1 might be interpreted as the result of a transformation which can be inverted to obtain a value in $[0, \infty)$ with the desired cardinal interpretation.

4.3 Inferring uncertainty from news-media textual data

Briefly put, our aggregate uncertainty measure will be the fraction of FT articles that contain one or more uncertainty keyphrases. To help make this precise, and to link it to U_{it}^*, let us introduce some formal notation. This will be slightly more general than is immediately necessary, to make it re-useable for the firm-level measure in Chapter 2, and to make clearer the ways in which this methodology could be nuanced and extended in future work.

In this Section we will stay at the level of theoretical (but in principle observable) quantities. Section 5.2 outlines how we map these theoretical observables to empirical observables.

We observe some textual corpus (e.g. the full text of all editions of the Financial Times published over the last 30 years), segmented into items that are dated by $t \in \mathcal{T}$ where $\mathcal{T}$ is the set of all observed time periods. In an attempt to extract a useful quantitative signal about latent uncertainty from this semi-structured mass of information, we will focus on three derived quantities:

- n_{it} is the number of items in our textual corpus that are dated to period t and are classified, by binary[5] subject classifier C_i, as referencing subject $i \subseteq \mathcal{I}$ where $\mathcal{I}$ is the set of all possible subjects
- $m_{it} \in \{0,1,\dots,n_{it}\}$ is the number of these items counted in n_{it} that are also classified by binary uncertainty classifier C_u as expressing uncertainty
- U_{it}, our news-media uncertainty measure, is the unitless scalar ratio

$$U_{it} \equiv \frac{m_{it}}{n_{it}} \tag{1}$$

Notice that U_{it} is a reduction of the full information available in (m_{it}, n_{it}). This reduction is parsimonious for the purposes in this Chapter, where n_{it} is relatively large, but we will need recourse to both m_{it} and n_{it} in Chapter 2 where n_{it} is smaller so that (m_{it}, n_{it}) conveys correspondingly less information.

U_{it} is an intuitively plausible observable counterpart to U_{it}^* with the following useful properties: it is normalised by news volume so will not exhibit variation simply due to variation in news volume; and, under the reasonable assumption that m_{it} and n_{it} are non-decreasing in U_{it}^*, U_{it} is non-decreasing in U_{it}^*. On the other hand, $U_{it} \in \left\{{}^0\!/_{n_{it}}, {}^1\!/_{n_{it}}, \dots, {}^{n_{it}}\!/_{n_{it}}\right\}$ is discretised whereas U_{it}^* is continuous, and there are an infinite number of non-decreasing functions that could map U_{it}^* to U_{it}.

Formally, inference about U_{it}^* from our observables can be obtained via maximum likelihood estimation or Bayesian methods. Both methods will require a statistical model $P_{m_{it},n_{it}|U_{it}^*}(m_{it}, n_{it}|U_{it}^*, \Omega_{it})$ of the true joint likelihood associated with the data generating process (DGP). Ω_{it} are any other observables that might be relevant to inference on (m_{it}, n_{it}).

Note that without loss of generality we can factorise the joint likelihood into the product of conditional and marginal likelihoods:

$$\begin{aligned} P_{m_{it},n_{it}|U_{it}^*}(m_{it}, n_{it}|U_{it}^*, \Omega_{it}) &= P_{m_{it}|n_{it},U_{it}^*,\Omega_{it}}(m_{it}|n_{it}, U_{it}^*, \Omega_{it})\, P_{n_{it}|U_{it}^*,\Omega_{it}}(n_{it}|U_{it}^*, \Omega_{it}) \\ &= P_{n_{it}|m_{it},U_{it}^*,\Omega_{it}}(n_{it}|m_{it}, U_{it}^*, \Omega_{it})\, P_{m_{it}|U_{it}^*,\Omega_{it}}(m_{it}|U_{it}^*, \Omega_{it}) \end{aligned} \tag{2}$$

From a purely statistical point of view this factorisation could be conducted in either order, corresponding to the two lines of (2). However, the factorisation in the first line of (2) has a natural interpretation in the following two-stage model.

> Assumption 1: the vector (m_{it}, n_{it}) is generated by a two-stage process. In the first stage, a realisation n_{it} is drawn of the corresponding random variable with conditional PMF $P_{n_{it}|U_{it}^*,\Omega_{it}}(n_{it}|U_{it}^*, \Omega_{it}) = P_1(n_{it}|U_{it}^*, \Omega_{it})$. (Loosely this might correspond to the editorial process of commissioning news articles.) In the second stage, a realisation m_{it} is drawn of

[5] More generally, classifiers could assign a score to each item.

the corresponding random variable, conditional on n_{it}, with conditional PMF $P_{m_{it}|n_{it},U_{it}^*,\Omega_{it}}(m_{it}|n_{it},U_{it}^*,\Omega_{it}) = P_2(m_{it}|n_{it},U_{it}^*,\Omega_{it})$. (Loosely this might correspond to the articles being written and expressions of uncertainty conveyed (or not) in the resulting text.) Thus we have:

$$P_{m_{it},n_{it}|U_{it}^*}(m_{it},n_{it}|U_{it}^*) = P_2(m_{it}|n_{it},U_{it}^*,\Omega_{it})\ P_1(n_{it}|U_{it}^*,\Omega_{it}) \tag{3}$$

Notice that n_{it} is, by assumption, already fixed at the second stage, so we can conduct inference on U_{it}^* based on (m_{it},n_{it}) and the conditional PMF $P_2(m_{it}|n_{it},U_{it}^*,\Omega_{it})$ assuming that n_{it} is fixed. However, this will not in general be efficient, since we discard any information about U_{it}^* that might be inferred from the first stage outcome n_{it}. That said, it is plausible that any dependence of n_{it} on U_{it}^* is weak, at least for the aggregate measures considered in this Chapter, since aggregate n_{it} is probably largely dictated by the relatively stable format of the Financial Times in terms of the size of a print edition and average article length. More importantly, limiting inference to information from the second stage frees us from the need to impose structure on the first stage.

To operationalise this empirically we need to impose more structure on the second stage.

> <u>Assumption 2a:</u> at the second stage, each of the n_{it} news items can be modelled as an <u>independent</u> Bernoulli trial, with success probability $p_{it} = U_{it}^*$ where 'success' means that the article expresses uncertainty[6]. The number of successes, m_{it}, is, by definition, binomially distributed, i.e. $m_{it} \sim Bin(n_{it},U_{it}^*)$.

We will discuss the rationale for and realism of this assumption below, but first let us note the implications. The conditional mean and variance of m_{it} are

$$\begin{aligned} m_{it} \sim \mathrm{Bin}(n_{it},U_{it}^*) \Rightarrow E_{m_{it}|n_{it},U_{it}^*}[m_{it}|n_{it},U_{it}^*] &= n_{it}U_{it}^*, Var_{m_{it}|n_{it},U_{it}^*}[m_{it}|n_{it},U_{it}^*] \\ &= n_{it}U_{it}^*(1-U_{it}^*) \\ E_{U_{it}|n_{it},U_{it}^*}[U_{it}|n_{it},U_{it}^*] &= E_{m_{it}/n_{it}|n_{it},U_{it}^*}\left[\frac{m_{it}}{n_{it}}|n_{it},U_{it}^*\right] \\ &= \frac{E_{m_{it}|n_{it},U_{it}^*}[m_{it}|n_{it},U_{it}^*]}{n_{it}} = U_{it}^* \\ Var_{U_{it}|n_{it},U_{it}^*}\ [U_{it}|n_{it},U_{it}^*] = Var_{m_{it}/n_{it}|n_{it},U_{it}^*}\left[\frac{m_{it}}{n_{it}}|n_{it},U_{it}^*\right] &= \frac{Var_{m_{it}|n_{it},U_{it}^*}[m_{it}|n_{it},U_{it}^*]}{n_{it}^2} \\ &= \frac{U_{it}^*(1-U_{it}^*)}{n_{it}} \to 0 \text{ as } n_{it} \to \infty \end{aligned} \tag{4}$$

By Chebyshev's inequality, unbiasedness and variance tending to zero are sufficient for consistency. Thus U_{it} is an unbiased and consistent[7] (in the large n_{it} sense) estimator of U_{it}^*. In fact it is a standard result that m_{it}/n_{it} is the maximum likelihood estimator (MLE) of p_{it}, or in other words U_{it} is the MLE of U_{it}^*. Therefore, by the general properties of MLE, we also have that U_{it} is an asymptotically efficient (within the class of estimators that condition on n_{it}) estimator of U_{it}^*. The estimation error $U_{it} - U_{it}^*$ is mean-zero binomially distributed, but by the de-Moivre–Laplace theorem this is well approximated by a normal distribution for large n_{it}, so that U_{it} can be thought

[6] In fact the weaker assumption of proportionately would suffice since the normalisation that we have imposed on U_{it}^*, i.e. $U_{it}^* \in [0,1]$ (see Section 4.2) is the same as the normalisation of p_{it}, by its definition as a probability measure, so that proportionality also implies equality: $p_{it} = U_{it}^*$.

[7] By Chebyshev's inequality, unbiasedness and variance tending to zero are sufficient for consistency.

of approximately as a proxy for U_{it}^*, subject to classical measurement error that becomes negligible as news volume becomes large.

If i encompasses multiple subjects and/or t encompasses multiple publication days then we might expect U_{it}^*, and thus the Bernoulli success probabilities, to vary between articles on different subjects or publication days. Retaining the assumption of independence between the Bernoulli draws corresponding to articles, we have U_{it} distributed as scaled Poisson binomial with mean and variance equal to the mean of the individual Bernoulli trials' means and variances respectively. If U_{it}^* is constant across all articles encompassed by i and t, then this simplifies to the binomial result above.

As an indication of the size of the measurement error, suppose that U_{it}^* is equal to 0.043, which is the sample mean of U_{it} at lower frequencies (see Section 6.1.3) where the noise-to-signal ratio should be low. Consider a typical publication day with 200 articles. Then according to the above model the standard deviation of U_{it} is 0.015, which is approximately one third of U_{it}'s sample mean and two thirds of its sample standard deviation. The latter figure we might interpret loosely as a noise-to-signal ratio. The sample mean can thus be distinguished from zero, but we would clearly want an errors-in-variables interpretation of U_{it} in a regression context at daily frequency. At lower frequencies, where we are aggregating over a greater number of articles, the noise-to-signal ratio is lower (e.g. 6.6% at monthly frequency assuming 5,000 articles; 2.0% at annual frequency assuming 60,000 articles) such that it might be more reasonable to neglect the measurement error.

The assumption that $p_{it} = U_{it}^*$ is strong. We have no particular reason to expect that p_{it} varies linearly with U_{it}^*. However, we do expect p_{it} to be monotonically increasing in U_{it}^*. This leads us to relax Assumption 2a as follows:

> <u>Assumption 2b:</u> as Assumption 2a except that $p_{it} = h(U_{it}^*)$, with $h(\cdot)$ a strictly increasing function.

Under Assumption 2b, U_{it} is an unbiased and consistent estimator of $h(U_{it}^*)$. If $h(\cdot)$ was known and invertible then we could estimate U_{it}^* by inverting $h(U_{it})$. In practice $h(\cdot)$ is unknown, so we lack a cardinal mapping between $\mathrm{E}[U_{it}]$ and U_{it}^*, but we know they have the same rank ordering. This partly motivates our consideration of rank correlations when comparing U_{it} and σ_{it}.

Relaxing the assumption one step further, it is conceivable that p_{it} depends on other factors (though we see no obvious candidates), but so long as the dynamics of and dependence on these factors are appropriately restricted, the ordinal correspondence between $\mathrm{E}[U_{it}]$ and U_{it}^* is preserved. For example, the following assumption would suffice:

> <u>Assumption 2c:</u> as Assumption 2b except that $p_{it} = g_1(\boldsymbol{\Omega_i}) + g_2(\boldsymbol{\Omega_i})h(U_{it}^*)$ where $\boldsymbol{\Omega_i}$ is a vector of time-invariant factors, and within the domain of $\boldsymbol{\Omega_i}$ we have $g_2(\boldsymbol{\Omega_i}) > 0$, and $g_1(\cdot)$ and $g_2(\cdot)$ are such that $p_{it} \in [0,1]$.

Notice that by treating U_{it} as commensurable across all periods t, we implicitly assume that the DGP is invariant across those periods, which in the current framework includes that $h(\cdot)$ is time-invariant. Concretely this would mean that the propensity of FT journalists to express uncertainty at any given level of latent uncertainty is time invariant, not changing with journalistic or social fashion. Given the long and conservative pedigree of the FT this seems plausible (which it might not for some UK tabloids for example).

The assumption of independence between Bernoulli trials in Assumptions 2a and 2b is very convenient analytically. Notice that we only need assume independence between articles on the same subject i published in the same period t. Nevertheless, this is not innocuous. It is easy to imagine causes of dependence in the tendency to express uncertainty, between articles concerning the same subject and published within the same time period, especially where periods encompass many publication days. For example, once uncertainty has been expressed in one news item it might become part of a standard narrative about that subject that tends to be repeated in subsequent news item, inducing a positive serial correlation between the success probabilities of successive Bernoulli trials. Alternatively, to the extent that the uncertainty is a focal part of the news item, this uncertainty would no longer be 'news' after that publication, and so may be less likely to be mentioned again in subsequent news items even if the uncertainty remains unchanged, inducing a negative serial correlation in success probabilities. However, absent a clear prior on what the mechanisms are likely to be, and given the substantial increase in complexity from modelling such a dependence structure (likely to require numerical simulation) we leave this to future work.

5 Empirical implementation

This Section outlines our empirical implementation of the measurement framework that was introduced in Section 4. Section 5.1 lists our data sources. Section 5.2 lays out the mapping from the theoretical observables in the measurement framework to empirical observables. The remaining Sections highlight key challenges to practical implementation. Section 5.3 documents the prevalence and patterns of duplication in the raw FT data stored in Factiva, which has been neglected in the literature to date. We propose and apply a de-duplication method, and discuss the impact on our uncertainty measure. Section 5.4 highlights the importance of normalising by news volume, rather than using raw article counts as in previous research on UK data.

5.1 Data sources

Daily observations of m_{it} and n_{it} for 1 January 1982 to 30 April 2014 are derived primarily from the Dow Jones Factiva database. To help in cleaning and de-duplicating the data, and quality checking the results, we also drew on alternative news archive services Nexis UK and Proquest ABI/Inform[8], and on the Gale FT Historical Archive ('Gale' from here on) which provides electronic facsimile copies of the daily London print edition of the FT.

5.2 Mapping to empirical observables

To operationalise U_{it} at daily frequency we assume the following mapping from the theoretical observables in Section 4.3 to empirical observables:

- Textual corpus: full text of the London print edition of the Financial Times (FT), the leading daily business news publication in the UK, with an audience and content focus particularly suitable for measuring the uncertainty of major decision makers about business, economic and financial matters (see Appendix A.1 for quantitative evidence). Using a single publication keeps the data collection burden manageable, and minimises the variation in structure and format which might otherwise cause variation in the correspondence between U_{it} and U^*_{it}.

[8] Factiva coverage of the FT nominally begins 1 January 1981 but coverage for 1981 is unreliable. Nexis UK coverage starts on 1 January 1982; Proquest on 31 May 1996.

- Demarcation of items: items are defined as unique FT articles. More generally items could be more granular (e.g. paragraphs, sentences, phrases) or less (e.g. whole sections of a daily edition).
- Dating of items: publication date (which is presumably within a day or two of authorship for most articles in a daily newspaper like the FT) as recorded in Factiva.
- Set of observed time periods, $\mathcal{T}$: a canonical set of FT publication days (see Appendix A.2).

The classifiers bear a more extended discussion below.

Temporal aggregation, to generate lower frequency observations, involves separately summing m_{it} and n_{it} over the publication days in the period, and calculating U_{it} from these (as opposed to averaging daily U_{it}).

5.2.1 Uncertainty classifier, C_u

We classify an item as expressing uncertainty if its full text (including headline) contains any of the keywords[9] "uncertain", "uncertainty", or "uncertainties", which for brevity we will collectively refer to using wildcard notation as "uncertain*". The extant literature is largely constrained to the first two of these keywords[10]. Adding "uncertainties" increases the number of uncertainty articles by 14.3% without materially changing the semantic range. The corresponding boost in the signal-to-noise ratio is particularly helpful when dealing with disaggregated uncertainty measures like our firm-level measure in Chapter 2, where the number of uncertainty articles for a given company-year is typically in the single digits.

Serendipitously, "uncertainty" and its derivatives are already self-negated so that further negation – which can be serious a problem when interpreting keyphrase counts in other contexts – would result in double negation, which is relatively rare, especially in professionally edited news copy. For example, "not uncertain*" appears in only twelve of nearly two million articles in our sample period[11]. In corroboration, Baker et al. (2013) found that, in a human audit of 4,300 articles mentioning "uncertain" or "uncertainty" and "economic" or "economy", "only 1.8% of articles about economic policy uncertainty discuss low or declining policy uncertainty". We did not attempt to identify qualifiers of degree (e.g. "moderate uncertainty" vs. "extreme uncertainty") though this is one obvious direction for future development.

Adding further keyphrases would involve a trade-off between increasing signal strength (and potentially rounding out gaps in the semantic range of "uncertain*") versus increasing noise, due to uses of those keyphrases in senses other than those corresponding to latent uncertainty. We constrain our main comparative analysis to "uncertain*", so that it speaks more directly to the extant literature, and to minimise the chances of unknowingly adding noise from keyphrases whose semantic correspondence to our latent uncertainty concept has yet to be established. However, refining this classifier is one obvious direction for future work, and we conduct some initial explorations in this direction in Section 8.1.

---

[9] Keyword counting approaches are a specialisation of the "bag of words" in the information retrieval literature.

[10] Haddow et al. (2013) count only "uncertainty". Alexopoulos & Cohen (2009) and Baker et al. (2013) also count only "uncertainty" and "uncertain". (Alexopoulos & Cohen refer to another paper of theirs that counted "risk" instead, but this is neither published nor available online.)

[11] Similarly, "no longer uncertain*" and "nor uncertain*" appear twice each; "never uncertain*" appears once. Of course there may be instances of negation in a phrasal rather than single word form, but these are likely relatively rare.

The language used to express uncertainty, and the semantics of "uncertain*" seem unlikely to have changed substantially over the sample period (unlike, for example, words closely related to the internet such as "surf" and "web"), so it seems reasonable to treat the results from this classifier, applied to articles from different periods, as commensurable.

5.2.2 Subject classifier, C_i

The specification of the subject classifier is the one fundamental difference between the aggregate uncertainty measure studied in this Chapter and the firm-level uncertainty measure in Chapter 2.

For the aggregate measure, i implicitly encompasses all subjects covered by the Financial Times, and n_{it} is set equal to the count of all FT articles.

The extant literature has included general national newspapers in the textual corpus, so that many articles discuss matters outside the scope of economic/financial/business matters that we are interested (e.g. "uncertainty" over the outcome of a sporting fixture or that evening's TV soap opera). The proposed solution has been to only count articles that also contain "economic" or "economy" (Alexopoulos & Cohen, 2009; Baker et al., 2013; Dendy et al., 2013)[12].

However, the FT has a tighter subject focus, so that among the 57% of articles containing "uncertain*" that do not also contain "economy" or "economic", we find that 95% contain other terms usually related to the economy (see Table 1)[13]. The 75% boost in "uncertain*" article counts from removing the requirement for "economic" or "economy" to appear, is particularly helpful when disaggregating to high frequency, or our firm-level measure in Chapter 2 where the number of articles per company-year is already typically in the single digits.

Table 1: Economy-related uncertainty articles not mentioning "economic" or "economy", 1984–2012

Among articles containing "uncertain*" but not "economy" or "economy", the percentage that contain the specified economy-related terms	
financ*	66%
bank*	34%
debt*	14%
credit*	15%
bond*	11%
equity | equities	16%
money*	18%
business*	38%
profit*	30%
earnings*	15%
revenue*	11%
wage*	2%
government*	36%
politic*	19%

[12] The most nuanced variant among these paper requires "econ*" to occur within five words of "uncertain*" (Dendy et al., 2013).

[13] Some of the remaining 5% may refer to other economy-related terms not included in our exploratory search. However, some likely come from non-economic news, for example in the Saturday lifestyle sections. Unfortunately, the 'section' field in Factiva is not reliably populated for of the sample, and the FT's own organisation into sections changes over time, so it is not feasible to use this to separate such sections. Calculating m_{it} and n_{it} for particular sections could be one direction for future work as better data becomes available.

policy*	13%
eurozone	2%
oil	10%
gas	6%
coal	2%
any of the above	**95%**
none of the above	5%

Notes: "*" denotes a wildcard. Percentages are calculated from un-de-duplicated record counts as obtained from Factiva search results.

5.3 De-duplication of news items

Duplication of items is a common problem in computational analysis of textual data. Where present, it can distort inference based on counts or frequencies of items (or ratios thereof) if left uncorrected. To the best of our knowledge no systematic analysis has been published to date regarding duplication of FT records in Factiva, despite these records being used in a growing body of research. This Section begins to address that gap.

Identifying duplicates requires some benchmark of what constitutes a unique article. This is not as straightforward as it might at first appear, as we will see below. However, for the sake of argument and concreteness, suppose that unique articles for a given day could be manually identified from the facsimile copy of the day's FT edition in Gale.

For a non-trivial fraction of these unique articles, there are multiple corresponding records (each ostensibly an 'article') in Factiva. Records that are identical to one another in all fields except the unique identifier (accessionNo) assigned by Factiva, are unambiguously duplicates of the same underlying FT article. However, most records that one would want to identify as duplicates exhibit some variation in the article text, such as:

- cosmetic variation: white space, punctuation, coding of non-alphanumeric characters such as currency symbols, and case (especially in the headline)
- headline formatting: for example with the section name (e.g. "COMPANIES AND MARKETS") pre-pended in one version and not in another
- substantive textual variations that a human reader would intuitively identify as different. These include localisation of articles for different regional audiences, since Factiva appears to contain articles from a mix of different editions on the same day. For example, in an article mentioning the UK Royal family, the US edition might include a brief explanation in the US edition of the identity and roles of leading Royal family members, that is not included in the UK edition due to assumed reader familiarity. Other such variations include corrections and editorial changes in later editions.

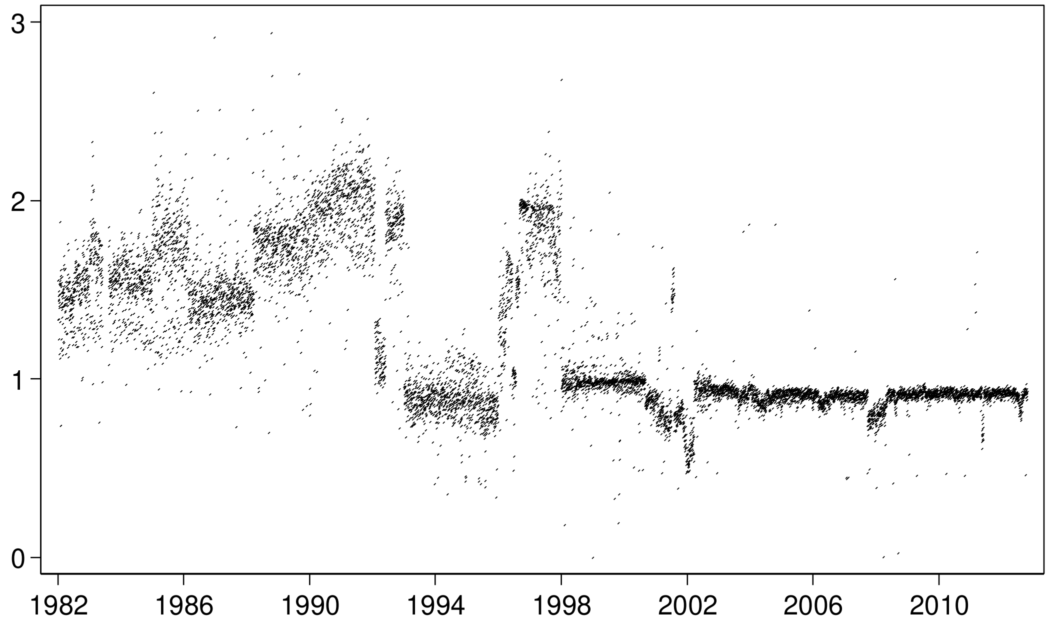
The degree of variation is on a continuum, and with the more substantive textual variations, two human readers might reasonably disagree on whether or not a given pair of articles constitutes duplicates of one another. Where to draw the line is ultimately a matter of judgement. Codifying intuitive human judgement into a rule that can be applied by machine is even more difficult, and is an open research question in the information retrieval literature.

We therefore designed a custom algorithm to identify likely duplicates based on relatively simple summary properties of Factiva records. We developed the algorithm iteratively, beginning with a simple version that looked for perfect duplicates (identical values in all fields except the unique Factiva database record identifier). This is discussed further below.

The key to developing a parsimonious algorithm to identify duplicates is to understand the nature of the variation that renders the corresponding database records non-identical. We targeted our diagnostic efforts on periods when the ratio of the total article count in Factiva to that in Nexis UK, shown in Figure 1 before any de-duplication, deviated substantially and persistently from unity. (We believe that Nexis UK contains relatively few duplicates for two reasons. First, its daily article count only changes slowly over the last 30 years, consistent with our prior expectation. Second, manual examination of all FT records in Nexis UK, for randomly selected days, revealed relatively few duplicates, as well as contents closely matching ProQuest and the facsimile copies in Gale. Thus Nexis UK seems unlikely to suffer from serious omissions. Indeed, in retrospect, though Factiva is more widely used in this nascent literature, it may have been preferable to use Nexis UK as our primary data source. However, this became apparent only after expending months of effort on data collection using Factiva.)

From Figure 1 it is clear that there are few duplicates since 26 June 2008, consistent with Factiva's claim, in its documentation, to have de-duplicated content added since that date. Our algorithm identifies only two duplicates (manually verified) beyond that date. Furthermore, duplication rates are mostly low beyond 2003. However, duplication rates are particularly erratic prior to 1993 and persistently high (around 2) during 1996-7.

Figure 1: Ratio of daily FT record counts in Factiva vs. Nexis UK, 1 January 1982–30 April 2014



Date axis tick marks are for 1 January in years labelled.
Ratio is missing for a single observation, 23 January 1999, due to a database error in Nexis UK.
Outliers truncated, numbering 31, of which 7 are for days with no articles in Nexis UK but non-zero article count in Factiva, implying an infinite ratio.

At each iteration of algorithm development we would focus on several days which had a ratio far from unity, manually checking for false positives in candidate duplicates and false negatives (residual duplicates) in the remaining records, and parsimoniously modifying the algorithm in an attempt to avoid these.

In addition to textual variations, duplicates also arise from the same article being tagged with different company codes by Factiva. Patterns in what appear to be date stamps embedded in the Factiva unique record identifier lead us to hypothesise that these duplicates arise from different versions of Factiva's tagging algorithm having being run at different times, without cleaning up the legacy copies.

Our final algorithm counts two articles as duplicates if all the following attributes are equal:

- publication date
- uncertainty keyphrase counts in each part of the article (headline, lead paragraph, tail paragraphs)
- headline *after* removing all non-alphanumeric characters (including punctuation and whitespace) and converting to uppercase
- (for publication dates up to 31 Dec 1988 only[14]) company codes applied by Factiva.

We believe this algorithm provides a reasonable working approximation for identifying duplicates, though it could undoubtedly could be improved in future work. In particular, our algorithm does not identify duplicates arising from the 'substantive textual variations' discussed above. Also, the deviations from a unit ratio in Figure 1 are partly due to differences between Factiva and Nexis UK, and probably also between different publication dates within Factiva (primarily comparing the 1980s with later years) in how a daily FT edition is split into 'articles' for database storage (see footnote 14 above for an example). These merit further investigation, in future research, to achieve a definition of an 'article' that is stable over the full sample period.

Note that we were only able to conduct de-duplication in this bottom-up manner on the subset of articles that contained uncertainty keyphrases and/or were tagged with a company code of a company in our firm panel from Chapter 2. This was because obtaining detailed article-level data on all of approximately 2million FT articles would have been prohibitively time consuming. As a result, we cannot calculate bottom-up de-duplicated counts for the aggregate denominator n_t, though we can for the aggregate numerator m_t, and the firm-level numerator m_{it} and denominator n_{it}. For n_t we use a top-down method, starting from total article counts in Nexis UK and applying the adjustments listed in Appendix A.3.

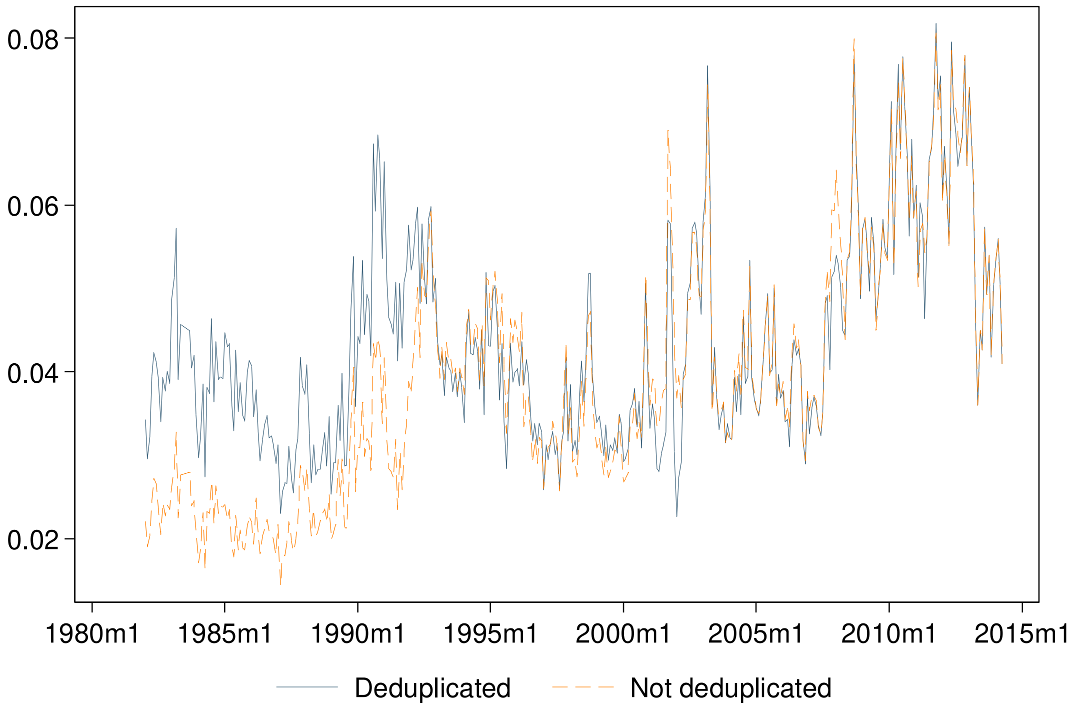
Based on the uncertainty articles counted in aggregate, m_t, and/or firm-level m_{it}, we estimate that on average 6.4% of Factiva FT records represent duplicates of unique FT articles, with this percentage being much higher in the periods with a high ratio in Figure 1. We retain a single record from each duplicate set, taking the union of company codes sets where these differ, and retaining the record with the larger number of text characters where this differs.

Duplication affects U_{it} only to the extent that it afflicts m_{it} and n_{it} differently (otherwise it cancels in the ratio). Unfortunately, manual investigation suggests that the duplication rate varies across articles even on the same day, especially before 1993, so that duplication rates on m_{it} and n_{it} may indeed differ. This does not completely wash out in the aggregate, as can be seen in Figure 2, which

[14] Summary articles about share stakes, appointments, and annual/interim report releases are split into separate items with identical headlines (e.g. "Share stakes") prior to this date, and requiring company codes to match prevents these being incorrectly identified as duplicates. Meanwhile, duplicates with different company codes appear to be more prevalent in the 1990s, so that removing this field from the duplicates criterion prevents those duplicates being missed.

compares aggregate uncertainty, U_t, when calculated from article data before and after deduplication. That said, the full sample correlation between the two versions is still high at 0.851. De-duplicating is likely to be even more important for disaggregated measures such as the firm-level measure developed in Chapter 2 of this thesis, where we are dealing with smaller news volumes, so that the signal-to-noise ratio is already relatively low, and (potentially non-Gaussian) noise from uncorrected duplication could loom larger.

Figure 2: U_t calculated on data before and after de-duplication, 1982m1–2014m4



5.4 Normalising by news volume

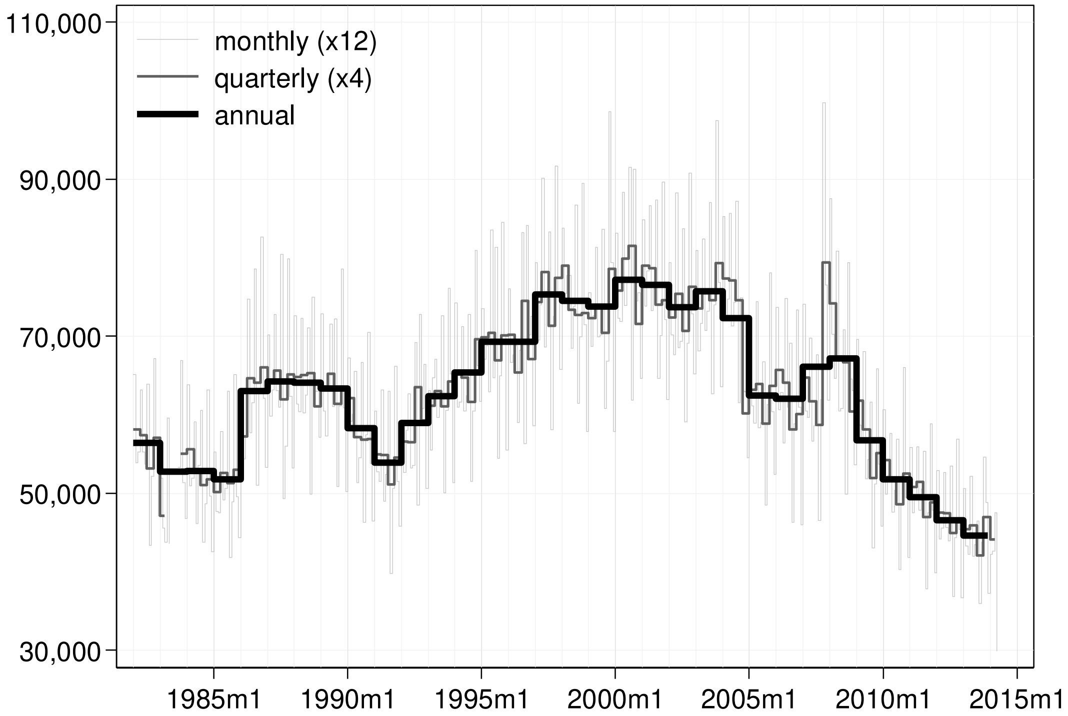
The nascent literature on news-media based measures of uncertainty has often used raw counts of articles expressing uncertainty, without normalising these by the underlying news volume. Alexopoulos & Cohen (2009) present their main results in terms of article counts without any normalisation[15]. Baker et al. (2013) normalise the news-media component of their US policy uncertainty measure by news volume, but cite unspecified technical barriers to doing so for their UK measure. Similarly, Haddow et al. (2013) use raw article counts for the UK.

In this Section we document both secular and high frequency variation in aggregate FT news volume, n_t. This presumably induces variation in m_t that is, at least in large part, unrelated to latent uncertainty. It will also affect disaggregate m_{it} and n_{it}. While the form of the normalisation that yields the best estimate of U_t^* and U_{it}^* will depend on if and how news volume is affected by latent uncertainty, some normalisation is likely to be better than none, and the simple division by total news volume in U_t and U_{it} seems a sensible first approximation.

[15] They show that their baseline macroeconomic VAR is not strongly sensitive to normalising the measure by news volume (their Appendix B) but do not report or otherwise analyse the normalised uncertainty measure.

Our dataset covers an estimated 1,998,165 unique articles of which 1,858,424 fall during our main 1984-2012 sample period. Figure 3 shows n_t at monthly, quarterly and annual frequencies.

Figure 3: Aggregate news volume (de-duplicated), n_t



Monthly news volume is relatively volatile in part due to variation in the number of publication days per calendar month. The remainder might be due to irregular publication of large supplementary sections (e.g. Special Reports) and random variation in the number of articles produced and the amount of space purchased by advertisers. Whatever the cause, failing to normalise by news volume would induce considerable noise in a news-media uncertainty measure at monthly frequency.

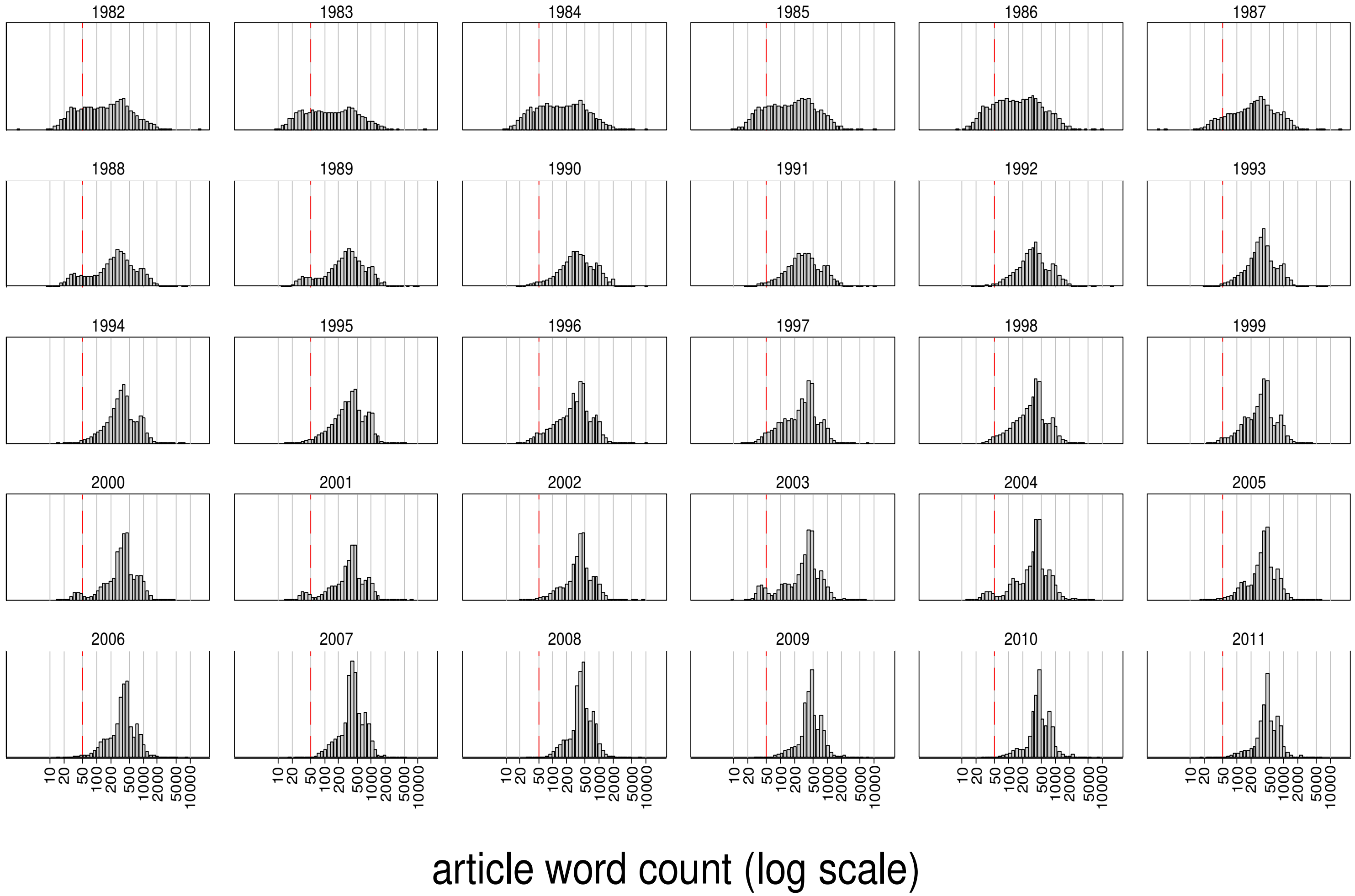
Quarterly news volume is much smoother, but exhibits a few large step changes. The reasons for the step up in 1986Q1/2 and down in 2004Q3 are unclear but these are replicated in the Nexis UK database and thus seem unlikely to be an artefact from Factiva or our deduplication algorithm. The two-quarter spike in 2008Q4/2009Q1 is likely driven by expanded news coverage at the onset of the Global Financial Crisis.

Secular trends are also clearly visible. The secular rise in the 1990s may reflect the phased launch of new international editions and the online edition. The secular decline in recent years probably partly reflects the steady shift in emphasis towards the online publication, including video and blog items not appearing in the print edition. Annual news volume varies by almost a factor two from the sample minimum of around 44,200 to the maximum of around 77,200 articles. Failing to normalise for this would induce spurious variation in U_t and U_{it} on a similar scale.

A potential additional driver of secular trends may be change in the distribution of the length of FT articles, as seen in Figure 4. It is not entirely clear the extent to which this reflects structural change in the FT's own article formatting, versus a change in the way that the facsimile copy is divided into records for storage in Factiva. For example, many of the short articles associated with the left tail of the distribution that is seen in the early 1980s but then recedes, refer to news snippets or corporate appointments announcements that are grouped under a larger heading in the facsimile copy and treated as a single record in later years[16]. Improving consistency over time of the way in which each edition is divided into 'articles' would be a substantial project beyond our current scope, but could be one direction for research. In any case, normalising article counts by news volume helps to control for such time-varying structural factors.

[16] Comparison of Factiva and Nexis results on selected days suggested that these article splitting practices differ between the two databases. This might contribute to the deviations from a unit ratio of Factiva and Nexis articles counts in Figure 2. However, it cannot be the primary cause of the deviations since these exhibit abrupt changes without correspondingly abrupt changes in Figure 4, and bottom-up analysis confirms that duplicates are substantially responsible (almost entirely so during 1996-7).

Figure 4: Distribution of article word count by year†



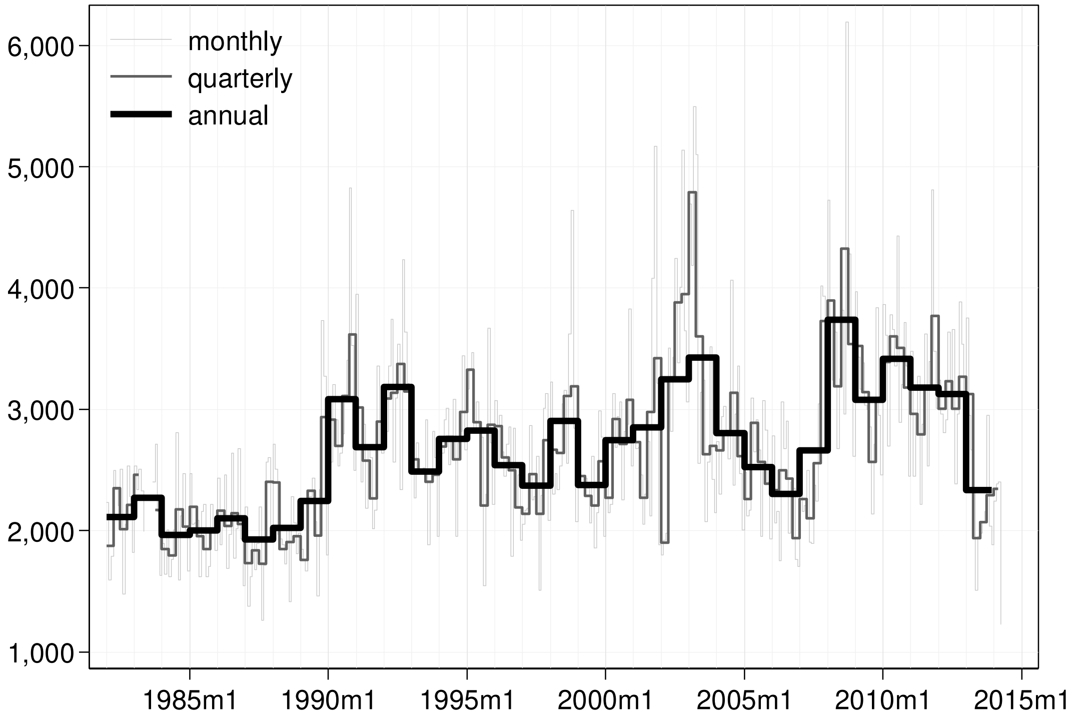
Notes: † Based on a de-duplicated subset of Factiva FT records (i.e. containing at least one uncertainty keyphrase and/or tagged with one of a target list of companies) for which we have word count metadata. We have no reason to expect that the distribution in the full population of unique FT articles differs systematically from that above. * Frequency is calculated per unit interval of log word count and displayed on a linear scale. Bins are defined identically across all years. Vertical red-dash lines mark word count of 50, an arbitrary threshold to categorise unusually 'short' articles.

The effect of failing to normalise by news volume at the aggregate level can be seen by comparing m_t and U_t in Figure 5. For example, at quarterly frequency uncertainty would be judged to be higher around the Iraq War in 2004Q1 than following the collapse of Lehman Brothers in 2008Q4, which we suggest does not accord with lived experience. At all frequencies the uncertainty ordering of the collapse of Lehman Brothers and the Eurozone crisis would be reversed.

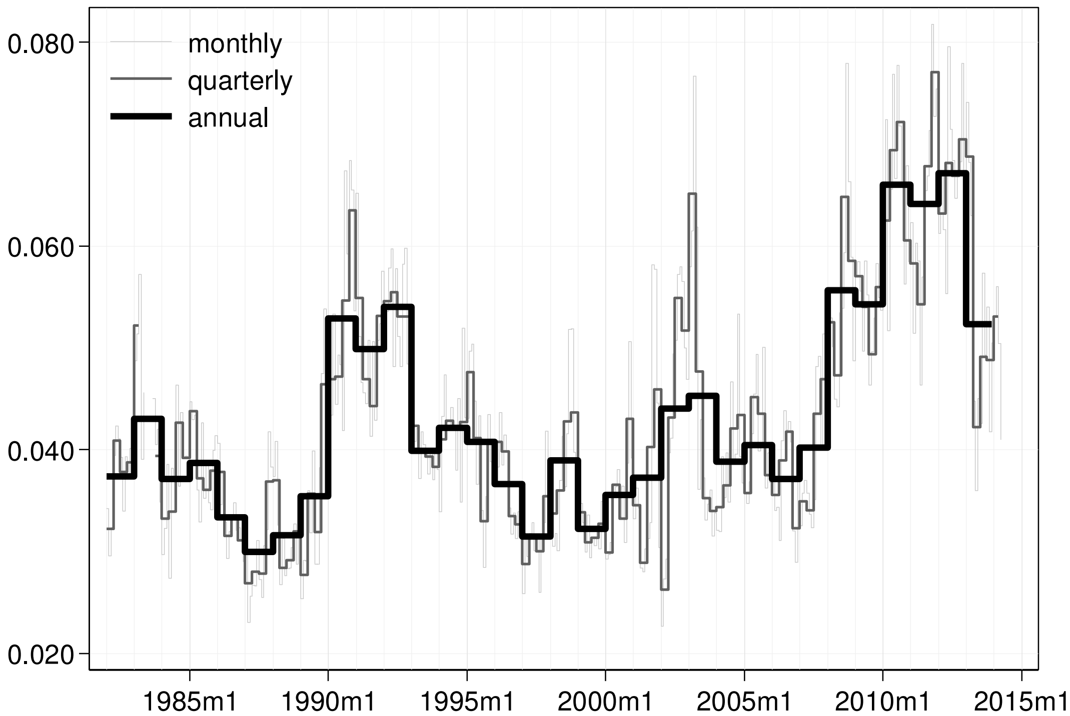
That said, some important features of U_t also appear in m_t, suggesting that these are not strongly sensitive to the specific normalisation: the timing of temporally local peaks is largely preserved; m_t remains elevated after 2010 just like U_t, showing that the continued elevation of U_t is not solely attributable to declining news volume; and the sign of first differences, i.e. the direction of movements, is little affected by low-frequency changes in n_t.

Figure 5: The effect of failing to normalise by news volume: m_t compared to U_t

m_t

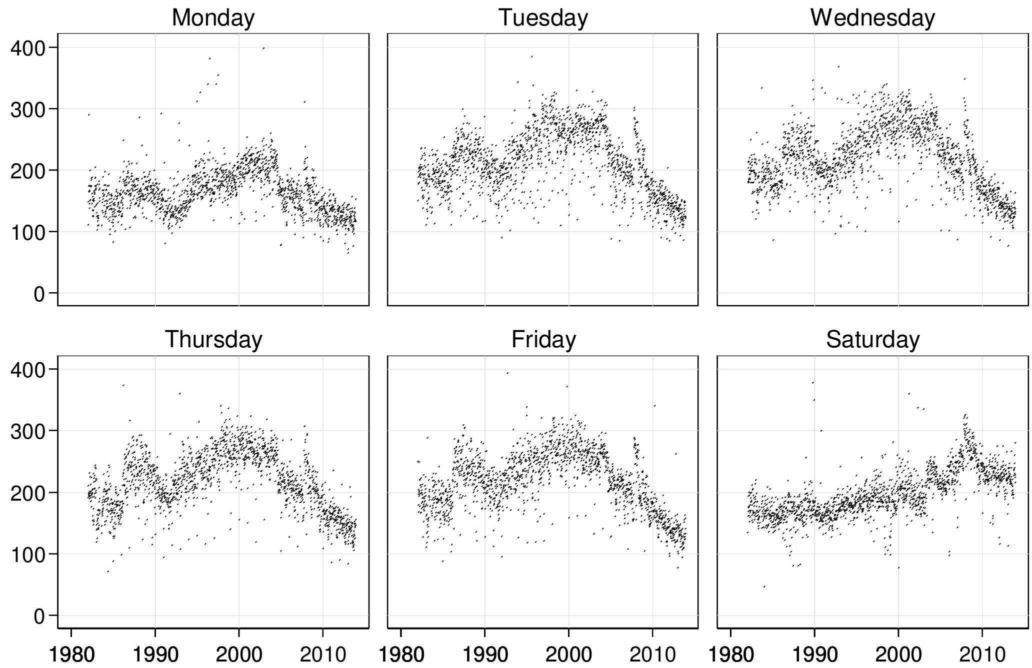


U_t



Finally, when it comes to analysis at daily frequency, there is significant day-of-week seasonality, as seen in Figure 6. For example, there are typically fewer articles on Mondays. Failing to normalise by news volume would thus induce substantial weekly seasonality that may not be present in U_{it}^{*}.

Figure 6: News volume, n_t, by day of week



Note: observations with more than 400 articles per days were truncated from the plot.

6 Comparative analysis

In this Section we compare aggregate news-media uncertainty, U, and stock returns volatility, σ.

Data preliminaries including sample selection, estimation of σ, and key properties of the data are summarised in Section 6.1. Our narrative analysis is presented in Section 6.2. Contemporaneous correlations are examined in detail in Section 6.3, with a focus on monthly frequency data. Higher frequency dynamic interdependencies are explored using Granger causality tests on a bivariate VAR in Section 6.4.

6.1 Data preliminaries

6.1.1 Sample selection

Our graphical and narrative analysis below considers all available data from 1981m1-2014m4. The FT coverage gap in all available electronic databases from 2 June to 8 August 1983 inclusive (see Appendix A.2) is reflected in missing values for 1983m6-1983m8 and 1983q2-1983q3, but we scale calendar year 1983 news volume and uncertainty measures pro-rata.

In quantitative analyses we focus on the longest span of complete calendar years (to ensure a common sample across all frequencies) for which we have continuous coverage (to avoid the complications of gaps in the time series) at the article level (to enable de-duplication; for 2013m11-2014m4 we only have top down counts from search results).

In common with the literature that compares stock volatility to non-financial variables (e.g. Campbell et al., 2001; Schwert, 1989b) we focus our narrative and correlation analyses primarily on monthly data, albeit with some consideration of other frequencies to fill out the picture. *A priori* we expect interactive dynamics between U and σ to occur primarily at higher frequency, so we investigate these using daily and weekly bivariate VARs.

6.1.2 Estimating σ

To estimate aggregate volatility, σ, we follow standard methodology from the literature.

We operationalise σ as returns volatility on a market index. For the UK the FTSE100 or the FTSE All Share are popular choices. We would prefer the latter for its broader coverage. However, our sample period begins in 1982 before these indices become available (3 January 1984 for FTSE100 and 31 December 1985 for FTSE All Share) so we instead use the Datastream Global Equity Indices Total UK Market index[17], which is very similar to the FTSE All Share index (Pearson's correlation coefficient between the returns is 0.994).

At weekly and lower frequency we estimate σ as the sample standard deviation of daily returns multiplied by an annualising factor $\sqrt{252}$, as originally proposed by Merton (1980) and in line with the subsequent literature (e.g. Campbell et al., 2001; Leahy & Whited, 1996).

At daily frequency we generate two estimates of σ. To obtain daily estimates on the full sample, for use in correlational analyses, we use the absolute daily return multiplied by the approximate normalising factor of $\sqrt{\pi/2}$ suggested in (Schwert, 1989a)[18] and then multiplied again by the same annualising factor as above. However, these are very noisy estimates of σ (the relatively low daily correlations reported in Table 6 are probably symptoms of this)[19] so do not provide a robust basis for inference in the VAR of Section 6.4. There we use the realized volatility (RV) estimates of Heber, Lunde, Shephard, & Sheppard (2009) which are based on intra-day tick data. The drawback is that these RV estimates refer to the FTSE100 index rather than the Datastream index used elsewhere in this Chapter, and are only available from 2000. That said, the FTSE100 accounts for over 80% of the market capitalisation of the Datastream index, and a 13 year span should give sufficiently precise estimates to identify any material Granger causality. The Pearson's correlation coefficient between the RV estimates and the daily absolute return over 2000–2012 was 0.451.

In principle, an ex-ante measure of volatility, such as options-implied volatility, might be preferred. However, options-implied volatility only dates back the late 1990s for UK data. In any case the time

[17] Datastream series TOTMKUK.

[18] See his footnote 4 where he attributes the multiplying factor to Dan Nelson. The factor arises because we are using absolute returns, rather than squared returns as used in the usual estimator of standard deviation, to obtain greater robustness to outliers. Such robustness is particularly desirable when we have only a single observation contributing to the volatility estimate. Assuming the observations are approximately normally distributed, the expectation of the absolute value is $\sqrt{2/\pi}$ times the expectation of the square root of the variance.

[19] The standard non-parametric estimate is the absolute daily return, i.e. based on a single observation.

profile of options-implied volatility and realised volatility are so similar as to be practically indistinguishable by eye.

6.1.3 Univariate distributions

Table 2 reports sample summary statistics for U and σ in levels and first differences (denoted by Δ[20]). The mean of U is around 4.3%, and the interquartile range is 3.4% to 5.1% at monthly frequency. The mean of σ is around 15%, and the interquartile range is 9.9% to 16.5%.

Both variables are approximately log-normally distributed (though $\sqrt{U}$ gives a better approximation to normality for the daily data used in Section 6.4). They exhibit lower variance at lower frequency. consistent with mean reversion, and higher excess kurtosis at higher frequency. First differences are closer to normally distributed, with zero mean consistent with stationarity (the skewness in $\Delta\sigma$ is primarily due to a few large positive outliers). Similarly, Table 3 shows that the distribution of positive vs. negative movements is equally balanced.

Table 2: Sample summary statistics, 1984–2012

Variable /	**frequency**	**Mean**	**Std. dev.**	**Percentiles**			**Skew-ness**	**Kurtosis**	**# of obs.**
				25th	**50th**	**75th**			
U	daily (ex.)	0.0443	0.0216	0.0289	0.0407	0.0559	0.972	4.500	7326
	daily	0.0436	0.0211	0.0286	0.0404	0.0550	0.963	4.558	8909
	weekly	0.0432	0.0142	0.0329	0.0402	0.0508	0.951	3.908	1508
	monthly	0.0432	0.0124	0.0336	0.0399	0.0510	0.895	3.165	348
	quarterly	0.0432	0.0117	0.0340	0.0410	0.0506	0.849	2.975	116
	annual	0.0431	0.0105	0.0366	0.0399	0.0499	0.972	2.927	29
σ	daily	0.148	0.151	0.050	0.109	0.195	3.361	26.079	7326
	weekly	0.139	0.096	0.081	0.115	0.165	3.346	23.738	1508
	monthly	0.147	0.082	0.099	0.124	0.165	2.834	15.218	348
	quarterly	0.151	0.075	0.104	0.132	0.170	2.461	11.283	116
	annual	0.157	0.062	0.120	0.142	0.177	1.372	5.146	29
ΔU	daily (ex.)	0.0000	0.0235	-0.0145	-0.0001	0.0146	0.068	3.782	7325
	daily	0.0000	0.0235	-0.0147	-0.0001	0.0146	0.071	3.902	8908
	weekly	0.0000	0.0107	-0.0064	0.0001	0.0067	0.159	4.623	1507
	monthly	0.0001	0.0070	-0.0042	0.0003	0.0044	0.128	4.067	347
	quarterly	0.0003	0.0068	-0.0034	0.0001	0.0049	-0.042	3.390	115
	annual	0.0011	0.0067	0.0034	0.0016	0.0036	0.462	3.642	28
$\Delta\sigma$	daily	0.000	0.184	-0.088	-0.000	0.088	0.038	11.485	7325
	weekly	-0.000	0.088	-0.041	0.001	0.039	0.974	23.603	1507
	monthly	-0.000	0.068	-0.029	0.002	0.023	1.736	16.419	347
	quarterly	-0.000	0.074	-0.034	-0.003	0.026	0.474	8.362	115
	annual	-0.000	0.070	-0.047	0.009	0.039	0.378	3.756	28

Notes: daily (ex.) excludes non-trading days as per the baseline daily VAR analysis in Section 6.4.

Table 3: Sample sign distribution of first differences, 1984–2012

Frequency	ΔU	$\Delta\sigma$	# of

[20] First differences are defined with respect to the basic periodicity at each frequency. For daily data this means FT publication days (which includes all trading days) or trading days where non-trading days are excluded.

	% of obs. by sign			Sign symmetry. p-value	% of obs. by sign			Sign symmetry. p-value	obs.
	<0	=0	>0		<0	=0	>0		
daily (ex.)	50.3	0.2	49.5	0.240	50.0	0	50.0	0.157	7325
daily (full)	50.1	0.2	49.7	0.171		n/a		n/a	8908
Weekly	49.8	0	50.2	0.797	49.7	0	50.3	0.701	1507
Monthly	48.7	0	51.3	0.560	52.5	0	47.6	0.326	347
Quarterly	48.7	0	51.3	0.781	53.9	0	46.1	0.395	115
Annual	43.9	0	57.1	0.451	46.4	0	53.6	0.706	28

Notes: see also notes to Table 2. Percentages may not sum to 100% across rows due to rounding. Column entitled 'Sign symmetry p-value' reports two-sided p-values for the null hypothesis that 50% of signs are strictly positive. P-values are obtained by normal bootstrap with 999 resamples over non-overlapping blocks spanning 2 calendar months (rounded up at non-monthly frequencies) (see Section 6.3.1.2). We also verified that 50% sign share was encompassed by 90% confidence intervals based on the bias-corrected and accelerated bootstrap of Efron (1987) (not reported for the sake of space).

6.1.4 Time-series properties

Both U and σ are strongly and significantly autocorrelated at a lag of one month, both in the full sample (see Table 4) and in most rolling windows (see Figure 7), though autocorrelation of U is more stable over time. U is more persistent than σ.

Table 4: Monthly autocorrelations, 1984–2012

	u	σ
ρ_1	0.83 ***	0.66 ***
ρ_2	0.76 ***	0.48 ***
ρ_3	0.70 ***	0.38 ***
ρ_4	0.64 ***	0.31 ***
ρ_6	0.59 ***	0.30 ***
ρ_9	0.53 ***	0.23 **
ρ_{12}	0.46 ***	0.16

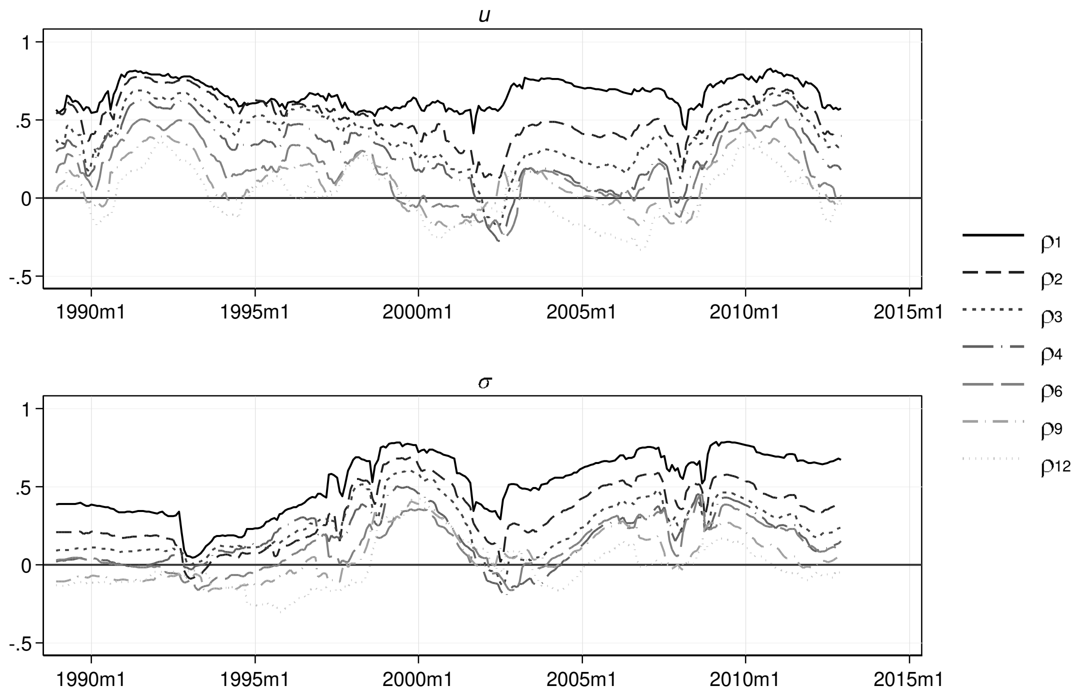
Notes: ρ_l is the sample autocorrelation coefficient at lag l. *, **, *** indicate significance at 10%, 5% and 1% levels respectively. Critical values are approximate, based on the assumption that the time series are Gaussian. Under the null that autocorrelation ρ_l equals zero (in which case the bootstrap is not required to account for serial dependence) ρ_l is normally distributed around zero with standard error calculated using Bartlett's formula for MA(l) processes.

How should we understand the greater persistence of U compared to σ? Suppose we can interpret the *impulses* in both U and σ as reflecting *impulses* in latent uncertainty. Then at least one of U and σ must have different persistence than latent uncertainty, and this would blur that measures accuracy as a proxy for short-run movements.

One can imagine σ settling *before* latent uncertainty if there is persistent uncertainty but the flow of new information has limited impact on the forecast distribution of returns, so that classically rational investors have no reason to modify their positions after adjustments to the initial uncertainty impulse have been made. The news media may still report the information along with uncertainty references, so that u does not settle prematurely. Alternatively, U might settle *after* latent uncertainty if news articles make retrospective references to past uncertainty. Distinguishing between these possibilities in future research might be achieved by, for example, classifying the textual uncertainty references into current vs. retrospective.

Higher order correlation is varies more over time. The half-life of U (the lag at which autocorrelation drops below 0.5) ranges from around one to six months. The half-life of σ is often less than a month though goes up to around four months in the late 1990s.

Figure 7: Autocorrelations of U and σ within a rolling 60-month window



Note: the sharp drop in σ autocorrelation in October/November 1992 is associated with the October 1987 stock market crash passing out of the rolling window.

No unit root behaviour is evident from eyeball examination of Figure 10 below. U eventually retreats from the sustained highs of the early 1990s and the recent financial crisis. The recent literature has generally modelled σ as fractionally co-integrated but covariance stationary. DF-GLS tests of Elliott et al. (1996) rejected the null of a unit root on daily data for 1984m1-2012m12 (see Appendix, pg. 68, Table 9). This rejection was robust to subsample analysis either side of the time points where the data was most suggestive of a potential break (e.g. at the onset of the recent financial crisis) and to change of frequency to monthly.

6.2 Narrative analysis

Our narrative analysis will focus on the monthly series. Dynamics related to events of obvious interest can be obscured at much lower frequency, and the amount of detail becomes unmanageable over multi-month spans at much higher frequency, although we will zoom in to weekly views for a few examples.

The selection of events is based on our subjective perception of what have been the more significant economic and financial events of the period. The event set labelled in Figure 8 and Figure 9 overlaps substantially with that used by Haddow et al. (2013) on UK data and Figure 1 of Bloom (2009) on US data. That said, as is natural for a UK dataset, we include more UK/EU focussed events and fewer US specific events than in Bloom (2009).

The association between the measures and narrative events is based on coincidence of timing rather than textual verification of the subject to which uncertainty references attach. A more exhaustive narrative analysis is beyond the scope of this paper but would be an interesting direction for future research[21].

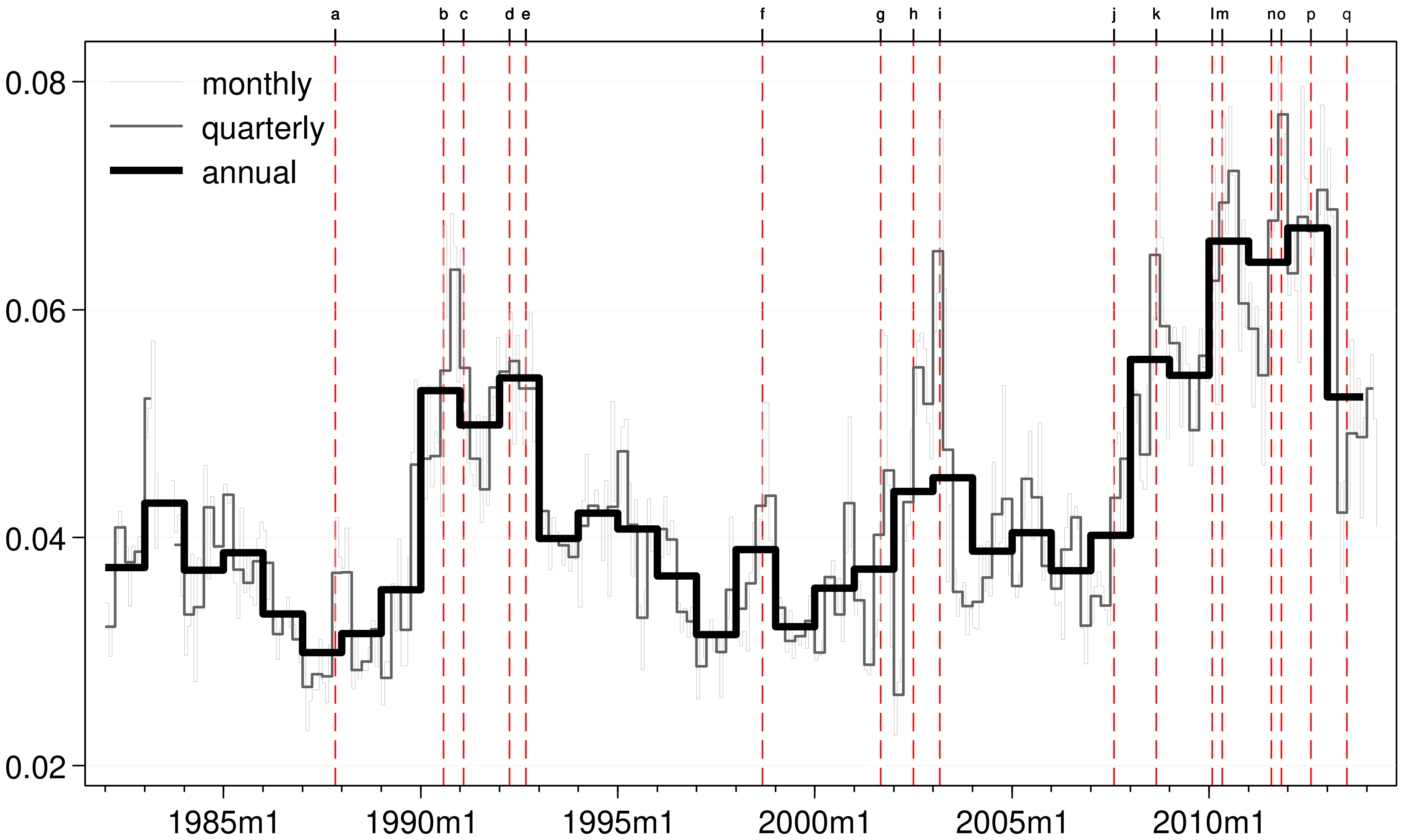
The discussion below seeks to illuminate the time-series plots of U and σ that follow. Figure 8 and Figure 9 show the variable separately with monthly, quarterly and annual frequencies overlaid. (Daily and weekly measures follow a similar profile but their higher volatility – partly measurement error – reduces the visual utility of their corresponding plots.) Stepped connecting lines help discern the correspondence between the overlapping periods at different frequencies. Figure 10 (levels) and Figure 13 (first differences) overlay the variables with separate plots for each frequency. Diagonal connecting lines help discern the co-movement (or otherwise) between U and σ. We also zoom in on selected events in weekly and monthly frequency plots that are interleaved with the discussion below.

To compare U and σ without applying any transformations implicitly assumes that the identity transformation is appropriate. Certainly it is conventional in the literature to interpret untransformed σ as (a proxy for) uncertainty. While there is little literature empirically *comparing* U and σ (indeed this is one of the gaps were are attempting to address), work that *combines* measures similar to our U and σ typically does so without applying transformations (e.g. Baker et al., 2013; Haddow et al., 2013), which implicitly assumes that the measures are comparable without transformation.

However, following the line of argument in Section 4.1, it is not clear what (monotonically increasing) transformation should be applied, absent a structural model tying both measures to an exogenously defined latent uncertainty concept, or some theory linking the measures directly. We therefore also present plots of the quantiles, which use only ordinal information, in Figure 12. Of course intuition and common practice suggests that the cardinal information is useful, so discarding it entirely would be a rather extreme response to this ambiguity about the appropriate transformation.

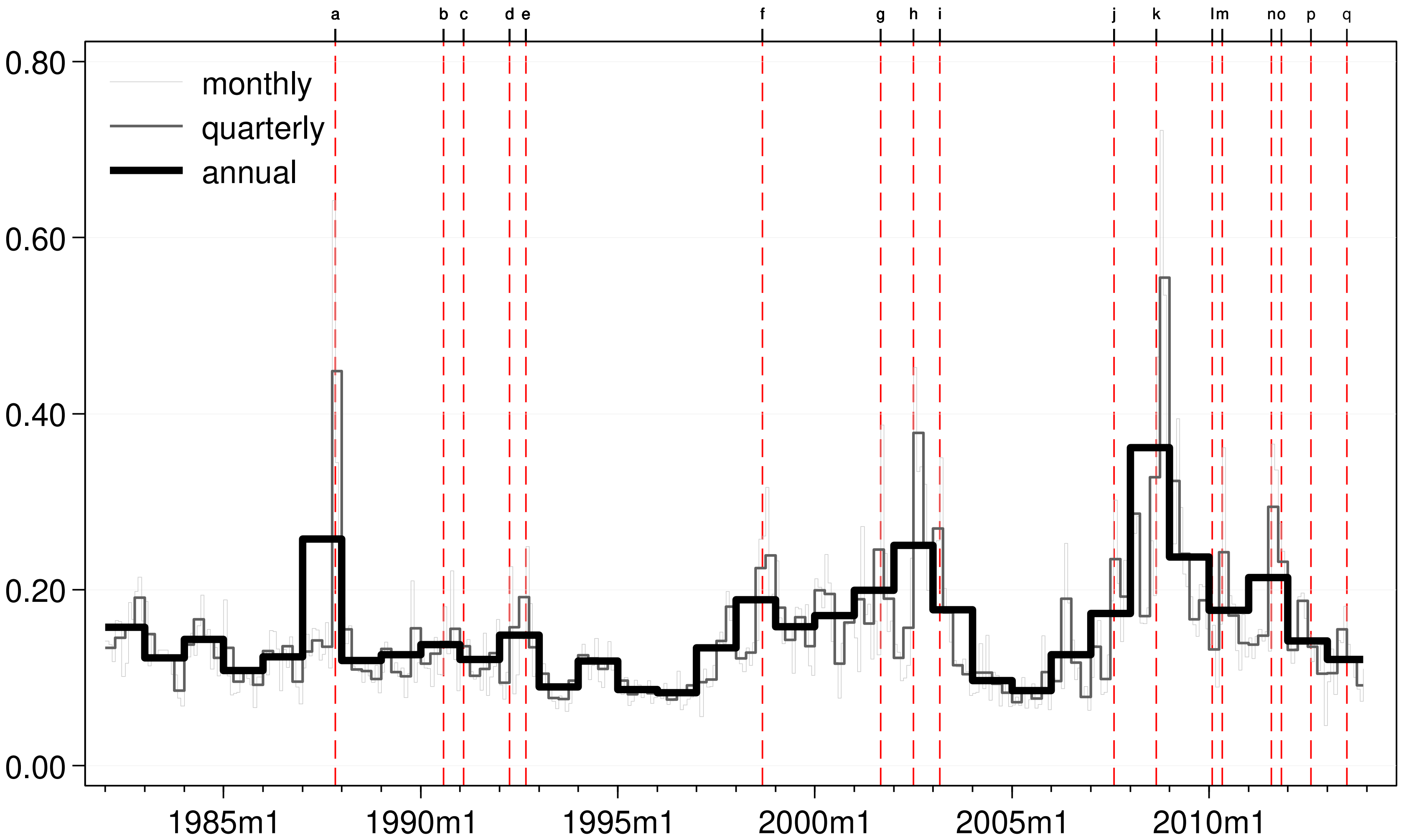
[21] One could examine the cross-section of uncertainty keyphrases with story-related keyphrases (e.g. "Greece", "bailout") in the same article (or paragraph or sentence etc.) and/or using article-level metadata on news subject from news database services.

Figure 8: Aggregate news-media uncertainty, U



Notes: See Table 2 on pg.23 for a description of the events associated with vertical red dashed lines. The annual measure is pro-rated for 1983 due to database gap for 2 June to 8 August; the corresponding months and quarters are set to missing.

Figure 9: Aggregate stock returns volatility, σ



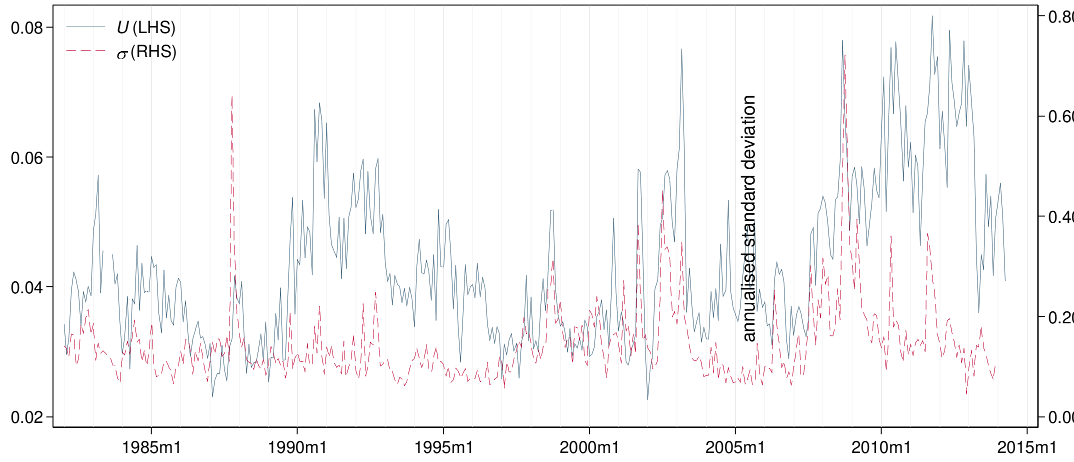
Notes: See Table 2 on pg.23 for a description of the events associated with vertical red dashed lines.

Table 5: List of events lines in Figure 8 and Figure 9

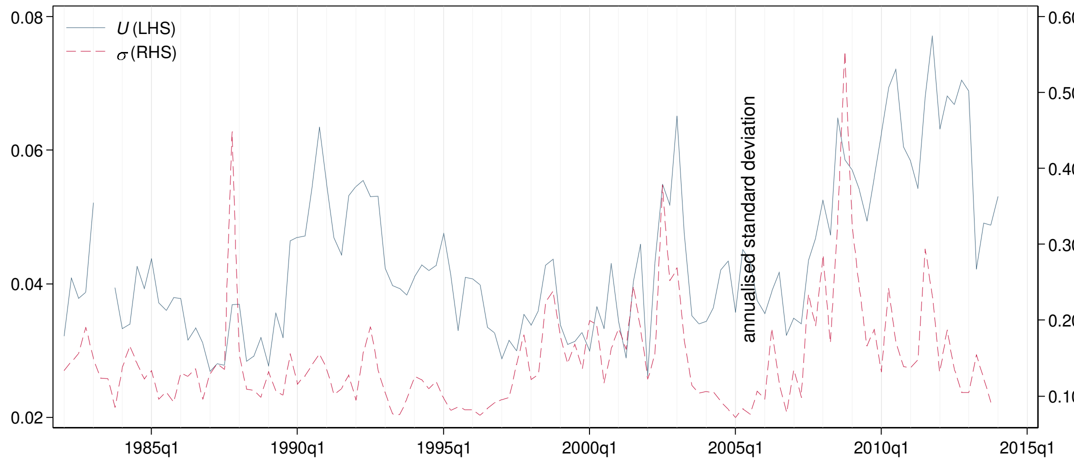
Label	Date	Description
a	19 Oct 1987	Black Monday stock market crash
b	2 Aug 1990	Gulf War: Operation Desert Shield starts
c	28 Feb 1991	Gulf War: Operation Desert Storm ends
d	9 Apr 1992	UK election
e	16 Sep 1992	UK exits European Exchange Rate Mechanism (ERM)
f	23 Sep 1998	Collapse of Long-Term Capital Management (LTCM) [dated to the recapitalisation]
g	11 Sep 2001	9/11 terrorist attacks in the US
h	Jul 2002	WorldCom accounting scandal
i	20 Mar 2003	Iraq War starts
j	9 Aug 2007	Subprime crisis onset: BNP Paribas freezes funds exposed to subprime mortgages
k	15 Sep 2008	Lehman Brothers bankruptcy
l	15 Mar 2010	Greek government debt crisis
m	12 May 2010	UK hung parliament [6-12 May]
n	2 Aug 2011	US government debt ceiling deadline
o	27 Oct 2011	EU bailout fund expanded to EUR1 trillion
p	26 Jul 2012	ECB governor Draghi promises to do ‘whatever it takes to save the Euro’
q	8 Jul 2013	French President Francois Hollande claims ‘crisis in the Eurozone is over’

Figure 10: U and σ overlaid, by frequency

Panel A: Monthly



Panel B: Quarterly



Panel C: Annual

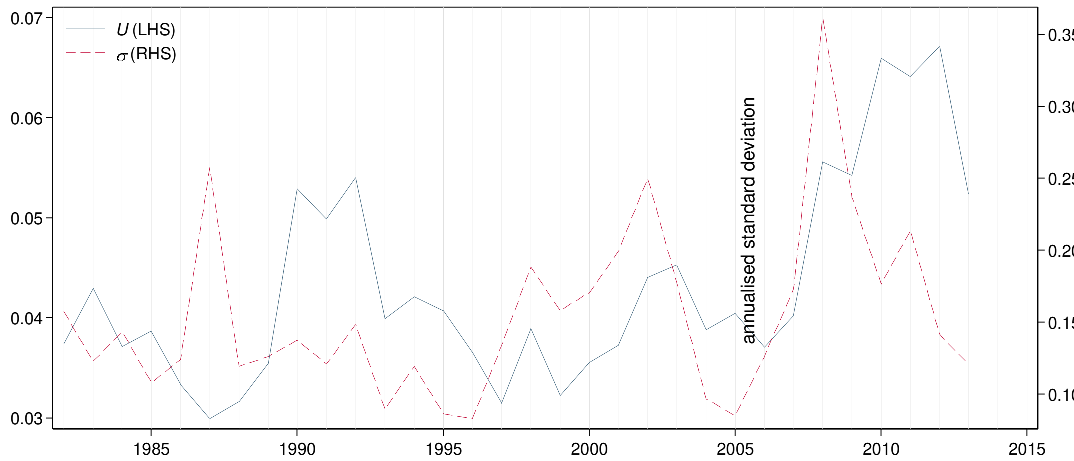


Figure 11: U and $\ln(\sigma)$ overlaid, by frequency

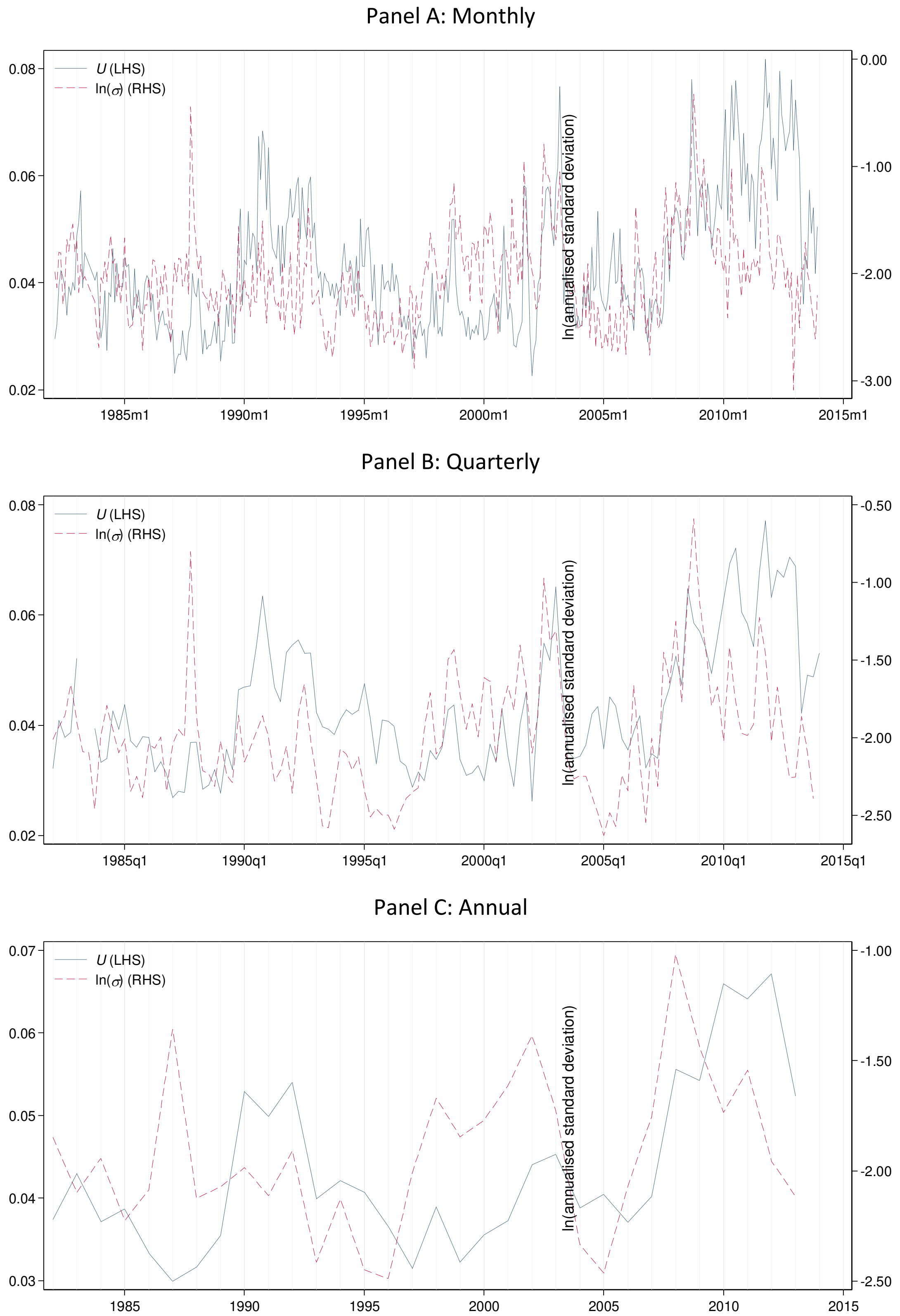
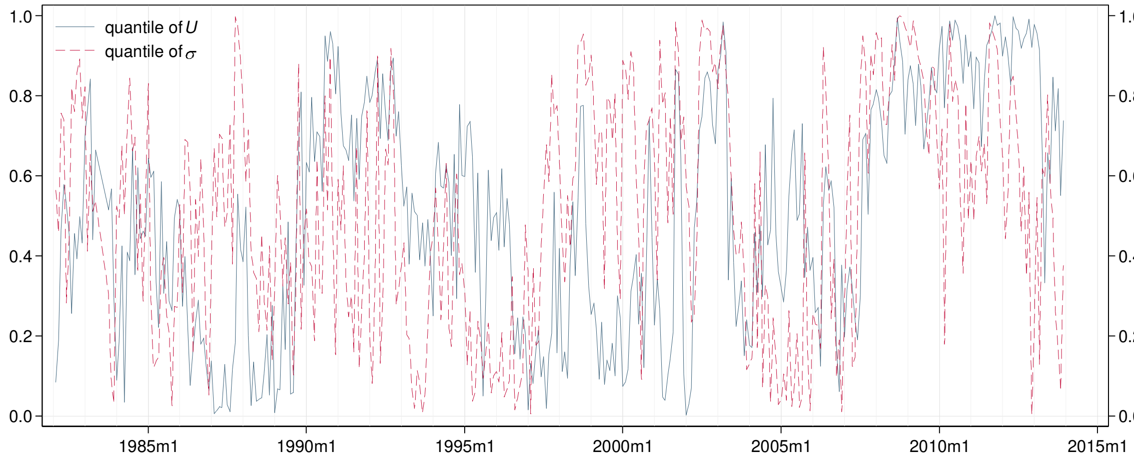
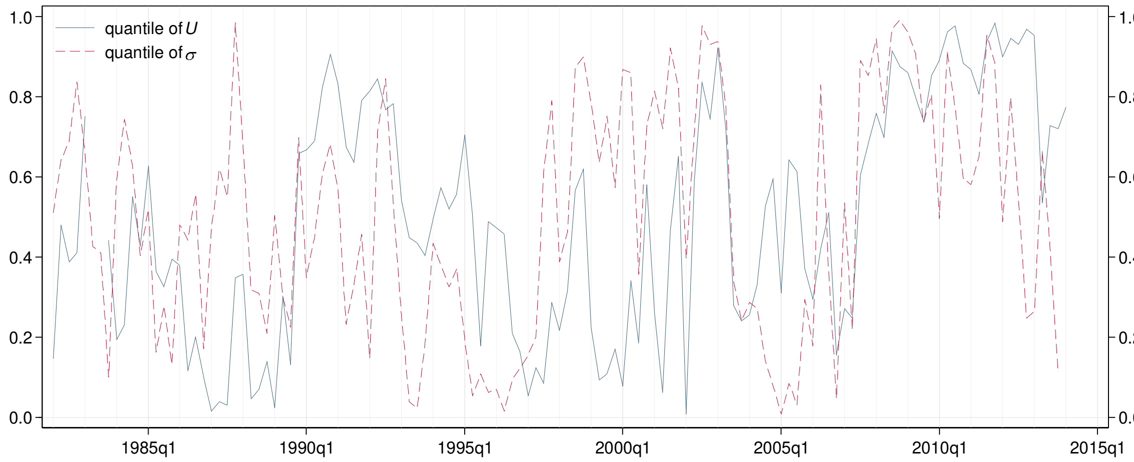


Figure 12: Quantiles of U and σ overlaid, by frequency

Panel A: Monthly



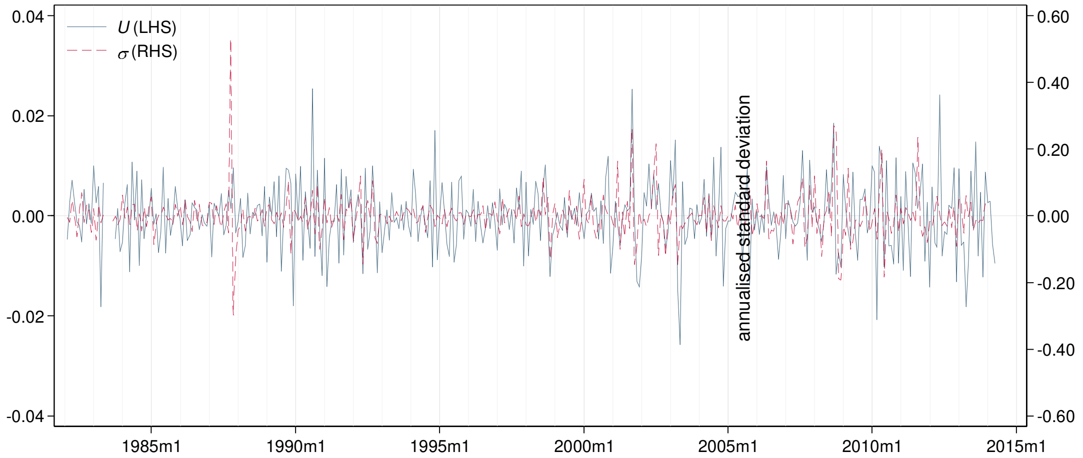
Panel B: Quarterly



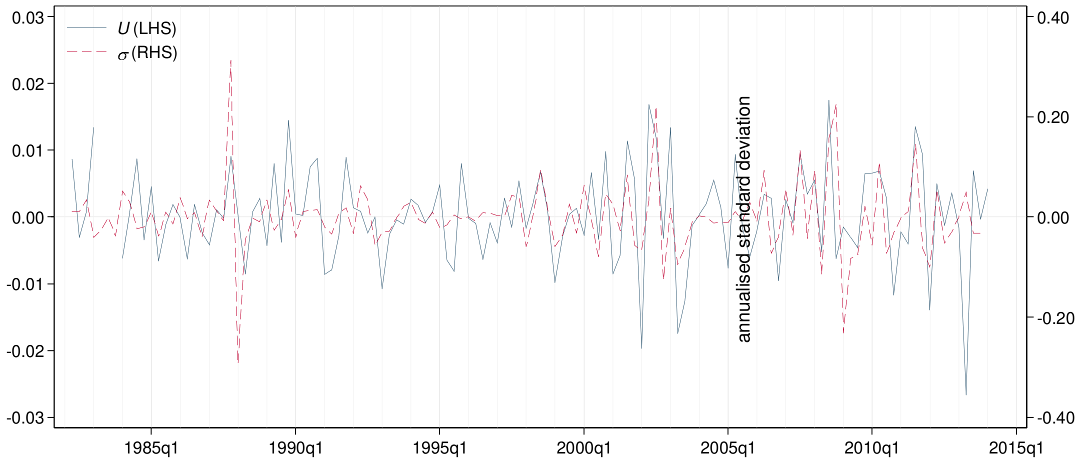
Panel C: Annual

Figure 13: First differences – aggregate news-media uncertainty, ΔU vs. stock returns volatility, $\Delta\sigma$

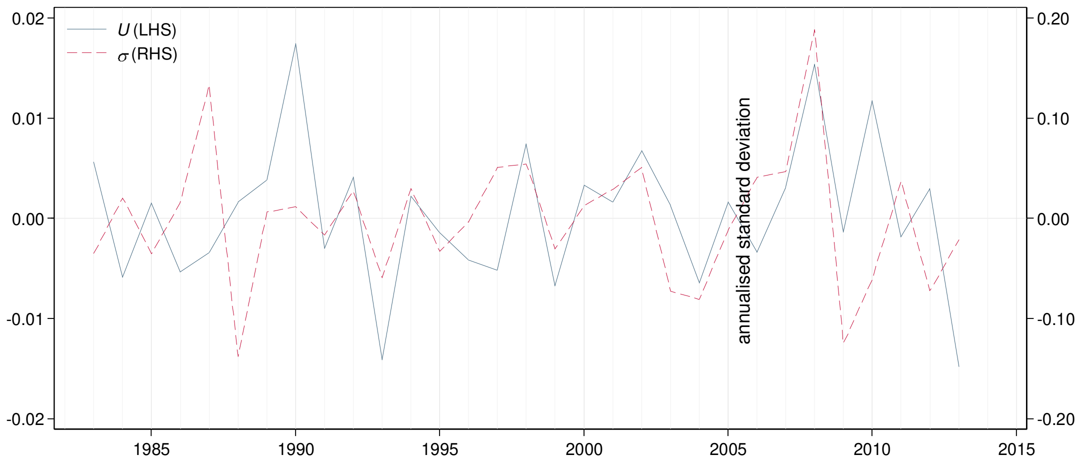
Panel A: Monthly



Panel B: Quarterly



Panel C: Annual

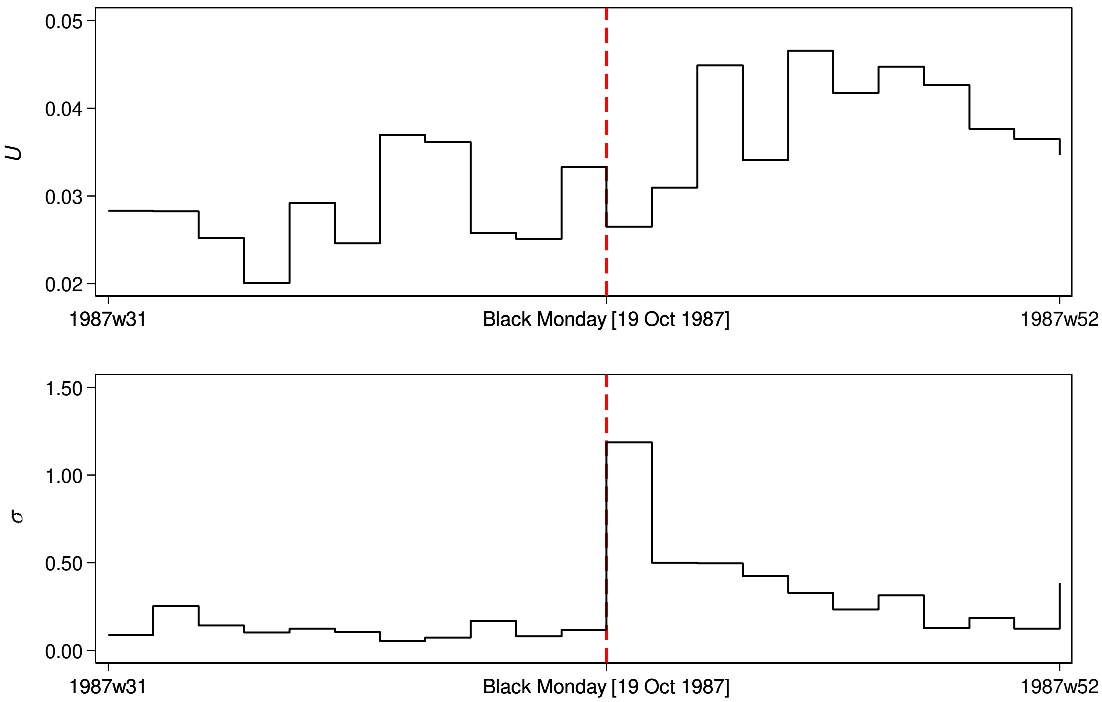


Let us start by noting two differences between U and σ that are apparent over the full sample.

First, the time profile of σ is dominated by large movements over short periods around major financial dislocations: the stock market crash of Black Monday (19 October 1987), and the collapse of Lehman Brothers (15 September 2008). U is elevated around these events too, but also reaches comparable levels at certain times in the first half of the 1990s, and in the years of the Global Financial Crisis (GFC) and Eurozone (EZ) crisis. We look at these periods in more detail below, but overall it appears that U may offer a broader-based view of uncertainty than σ. Naturally, this conclusion is tempered if we apply a skew-reducing transform to σ, such as $\ln(\sigma)$ (Figure 11), or if we consider only ordinal information (Figure 12).

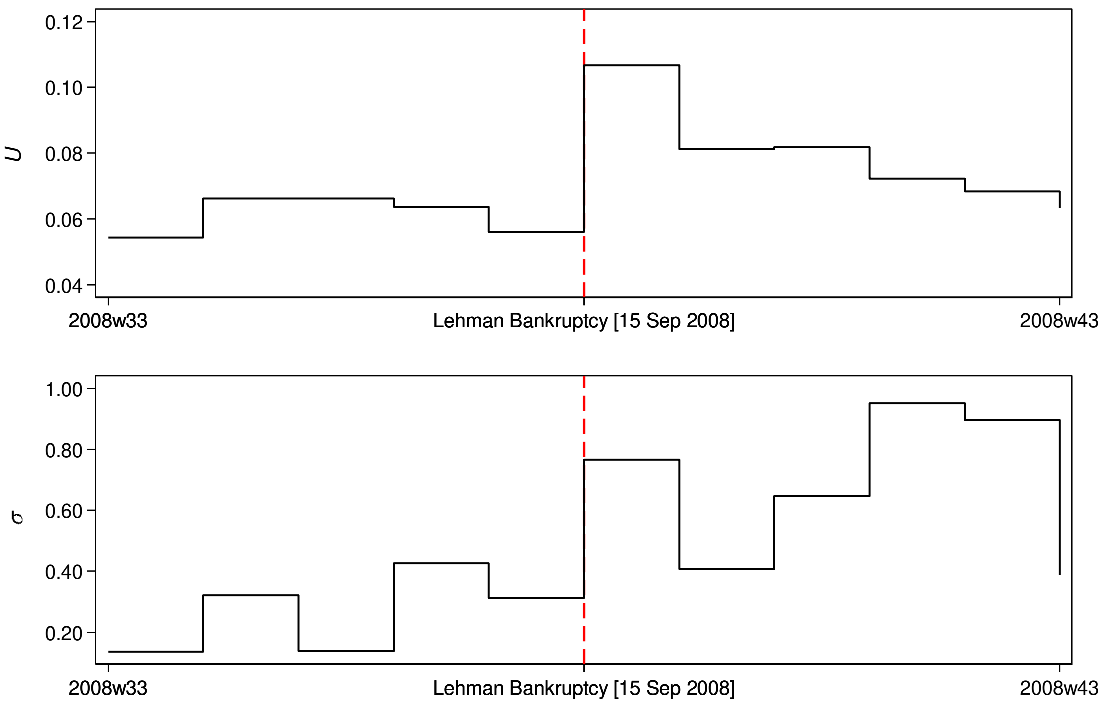
Second, the measures respond differently to major financial dislocations. Black Monday is associated with the largest positive monthly and quarterly increase in σ, to reach its second highest level in our sample. By comparison, the response of U is modest, similar to the first principal component of the battery of uncertainty proxies in Haddow et al., 2013. Here, U may better reflect the level of uncertainty in the real economy, since transmission of the stock market dislocation to the real economy was relatively muted – indeed the 1980s ended in the so-called Lawson boom. The two week lag in the rise in U is atypical. Usually U responds more promptly to major narrative events, as we will see below.

'Black Monday' – 19 October 1987
Weekly



The collapse of Lehman Brothers coincides with large spikes in both U and σ to near their sample maxima. At monthly frequency the local peak (also the sample maximum) of σ is delayed until following month (October 2008). This is at least in part because, with the emergency market closures, fewer than half of the trading days in September fell after the Lehman collapse. By contrast, U is able to reflect the spike in uncertainty more promptly because the newspapers kept publishing during this period. However, σ then drops back sharply within a few months and continues to trend down through the end of the sample, whereas U remains elevated until 2013q1, consistent with the conventional view that economic uncertainty remained elevated for years after the onset of the crisis.

Lehman Bankruptcy – 15 September 2008
Weekly

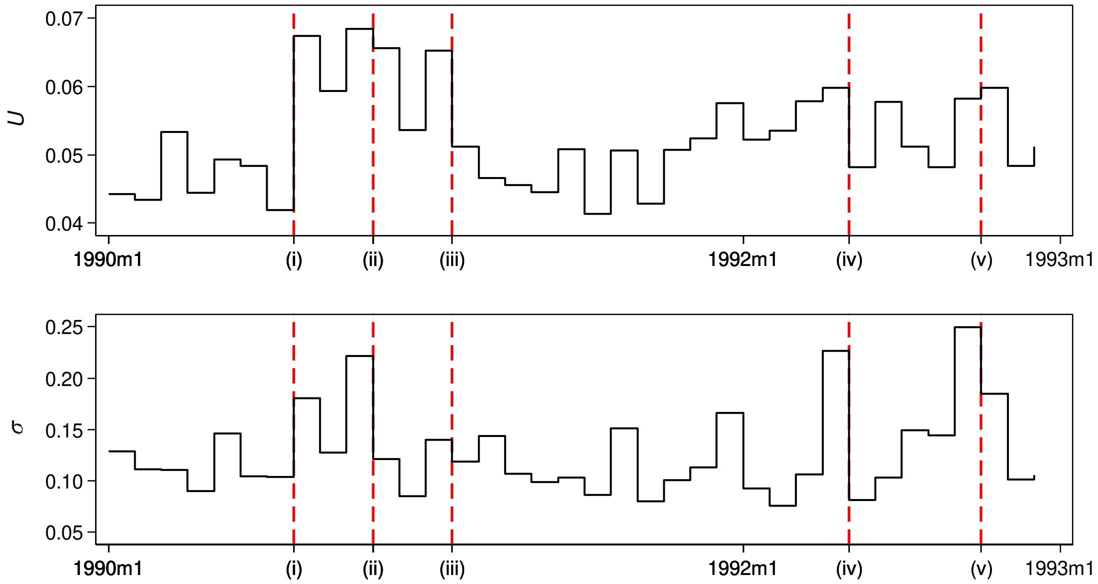


Arguably then, U is a more convincing indicator of the level of latent uncertainty during and following financial crises, even if the reason for the lag in response after Black Monday is unclear.

Let us now zoom in on a few periods that exhibit particularly substantial movement in U and/or σ.

In the early 1990s we see a sustained rise in U during the Gulf War (2 August 1990 to 28 Feburary 1991). σ rose at the start of the war but this higher level was not sustained. The rise in November 1990 coincides with the leadership battle in the ruling Conservative Party, leading to substantial political uncertainty and ultimately the ouster of Prime Minister Thatcher of 28 November 1990. Both measures show a drop off in uncertainty after the national election of 9 April 1992, and a rise coincident with the crisis around the UK's membership of the European Exchange Rate Mechanism (ERM) which was resolved by a forced exit from the mechanism in 16 September 1992.

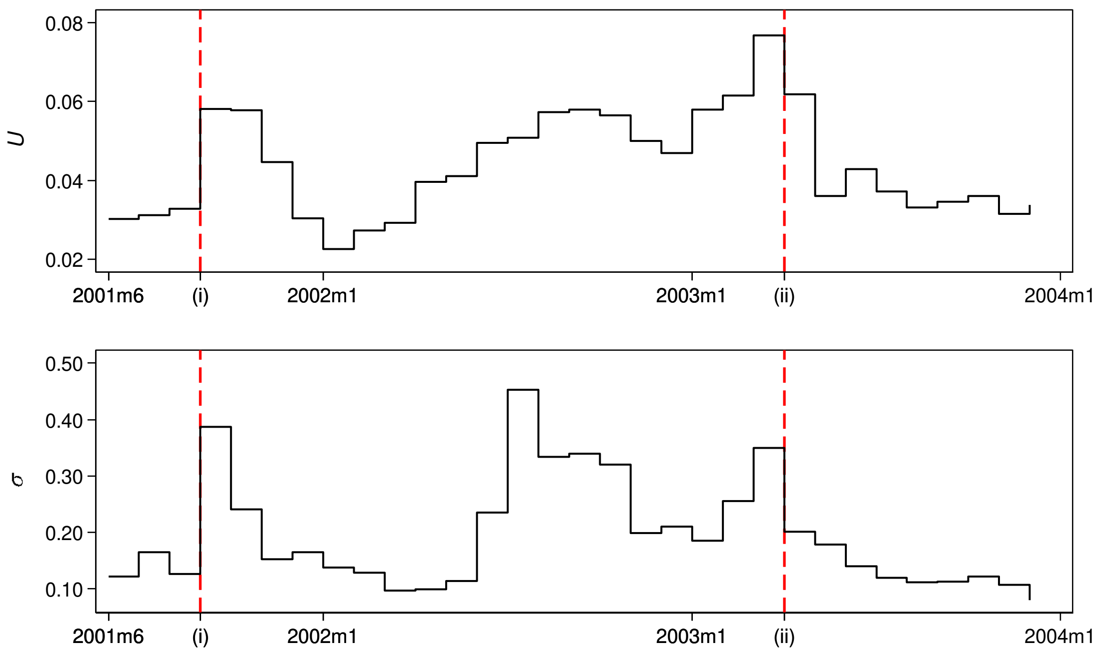
Early 1990s
Monthly



(i) Gulf War begins [2 Aug 1990]
(ii) Ouster of Prime Minister Thatcher [28 Nov 1990]
(iii) Gulf War ends [28 Feb 1991]
(iv) UK election [9 Apr 1992]
(v) UK exits ERM [16 Sep 1992]

The terrorist attacks in the US on Tuesday, 9 September 2001 ("9/11") and the subsequent war in Iraq are associated with clear peaks in uncertainty.

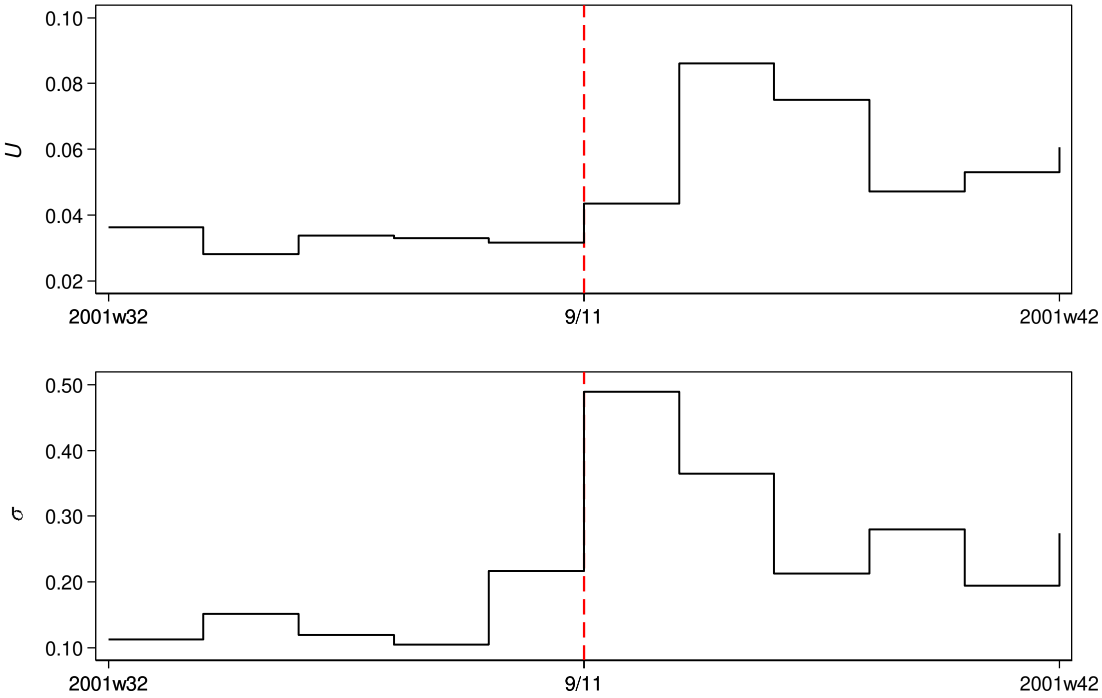
2001-3
Monthly



(i) Terrorist attacks in US [9 Sep 2001]
(ii) Iraq War begins (20 Mar 2003)

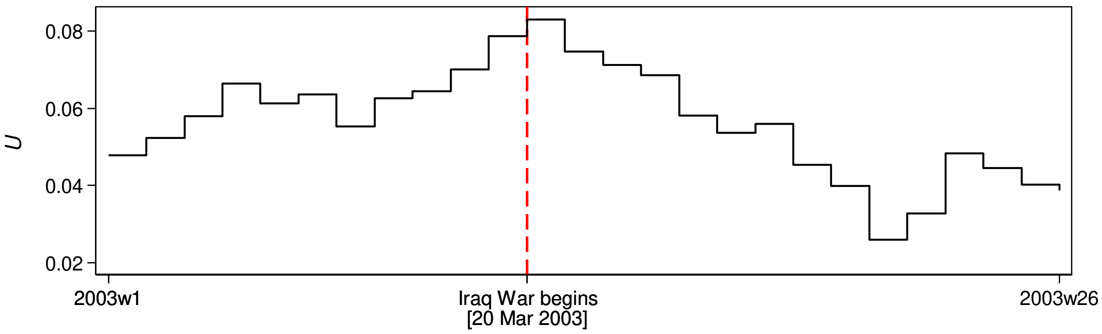
9/11 is the quintessential shock event. It is associated with a correspondingly sharp jump in both U and σ, although at weekly frequency the full jump in U is delayed by a week, at least in part because two of the week's six print editions pre-dated the attacks (of course the same is true for the daily returns from which σ is estimated, but the spike in volatility in the second part of the week sufficed to create a peak for the week overall).

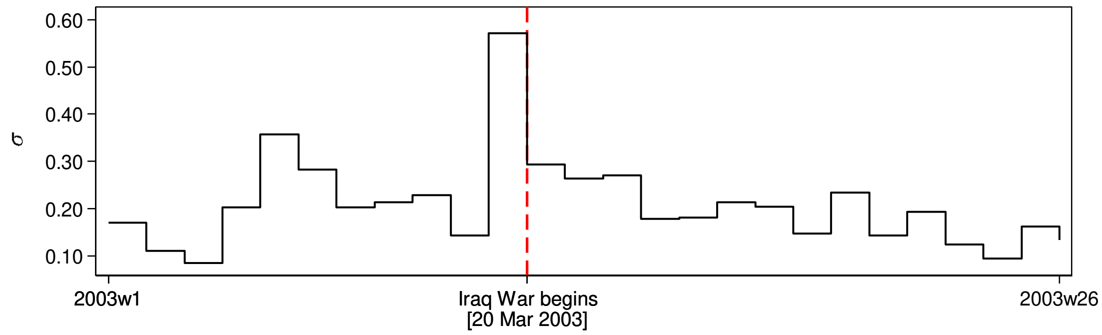
9/11



The Iraq War, by comparison, came after a multi-month build-up during which there was great uncertainty as to whether and when the war would commence, and which nations would join the US. This can be seen in both U and σ. The ramp up through 2002 may also be associated with the ongoing war in Afghanistan. When the US and its allies did finally attack on 20 Mar 2003, they toppled the incumbent regime in days, and much of the pent up uncertainty dissipated, though this occurs more gradually in U than in σ.

Iraq War
Weekly

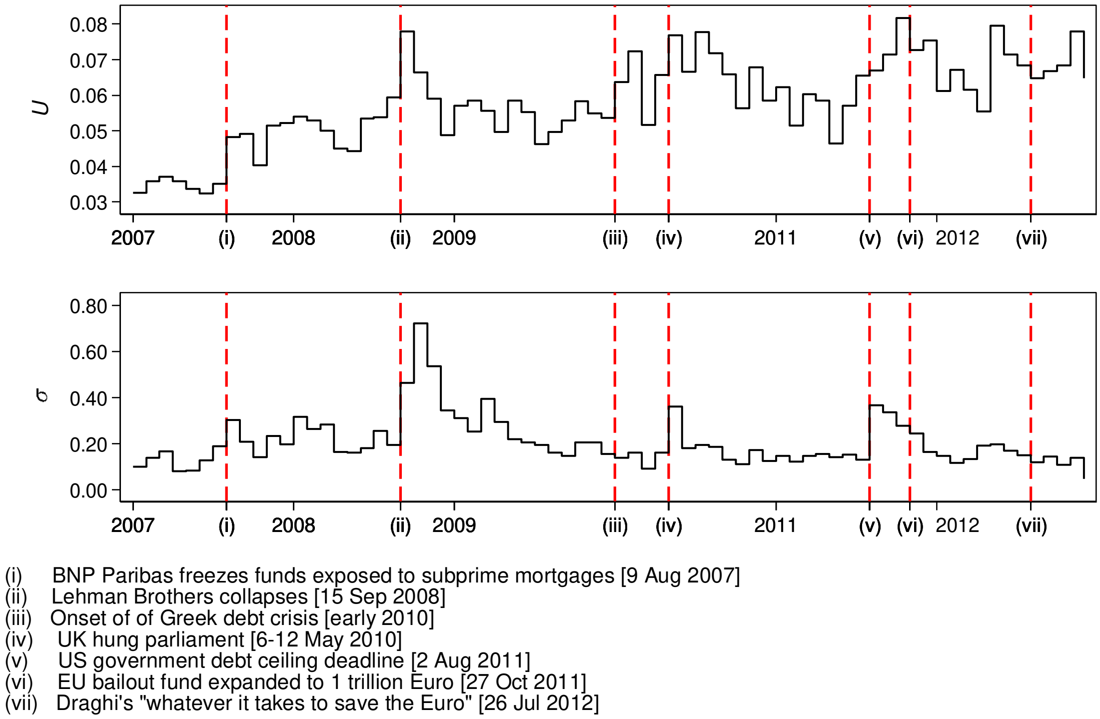




Moving on to the several years spanned by the Global Financial Crisis (GFC) and Eurozone (EZ) Crisis, we can see structure in U and σ consistent with conventional narrative accounts[22]. Both measures exhibit an uptick coincident with the event conventionally used to date the start of the GFC: BNP Paribas freezing three of its funds, which contained CDOs with exposure to subprime mortgages, on 9 August 2007. As discussed above, the collapse of Lehman Brothers stands out clearly in both measures.

Global Financial Crisis & Eurozone Crisis
Monthly

[22] The Guardian's (2014) Eurozone crisis timeline was particularly helpful in this analysis.



However, there are also clearly differences between U and σ in the post-Lehman period, beyond simply the suppression in the level of σ noted above.

Much of the subsequent variation in U can be linked circumstantially to events in the EZ crisis, or other key events. In conventional accounts the EZ crisis began with the Greek debt crisis in early 2010, at which time U rises but σ remains relatively muted. The sharp fall in U in 2013Q2 (see Figure 8) appears to be associated with a period of relative calm such that the French President felt able to claim in a speech on 8 June 2013 that "the crisis in the Eurozone is over"[23], even if this may have been premature in retrospect.

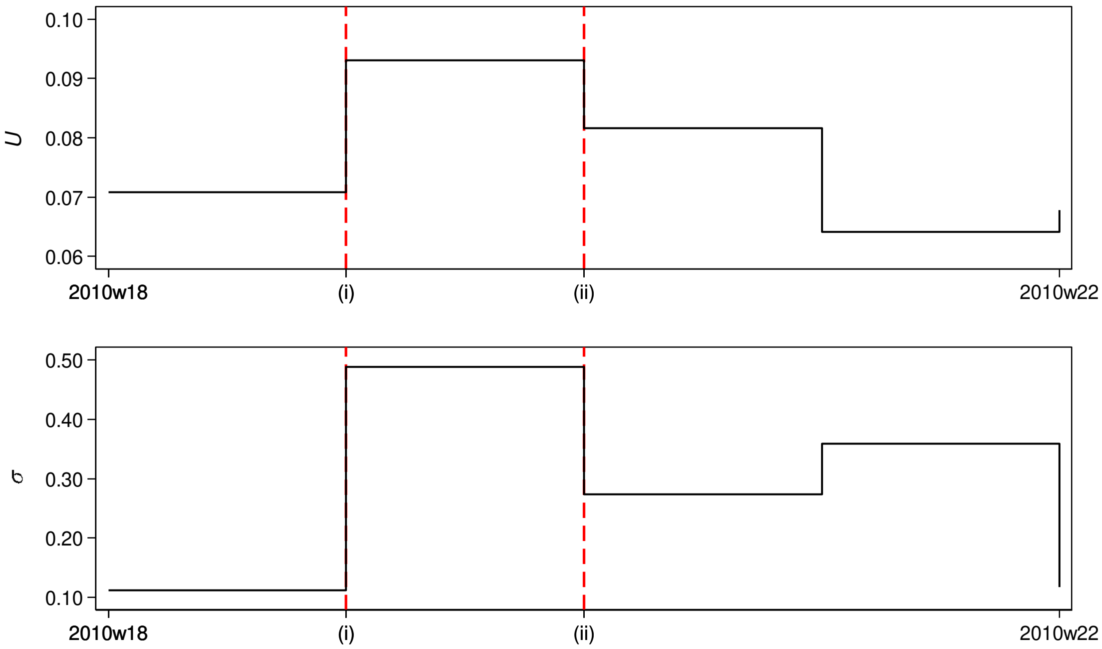
Significant uncertainty-related events in the post-Lehman period are so numerous, drawn out, and overlapping in time that persuasively identifying the relationship between all substantial movements in U and particular events is a demanding task beyond the present scope. However, we offer a couple of examples by way of illustration.

First, the national elections in the UK on 6 May 2010 led to a hung parliament for the first time since 1974, creating a great deal of uncertainty as to who would end up governing the country and setting policy. In the week between the election, and the formation of a coalition government on 12 May which dispelled this uncertainty, both U and σ rose sharply to around their 99th percentiles (see Figure below). The most obvious EZ crisis event that might be conflated in the dynamics here is the agreement on 10 May of EZ governments to a EUR500trillion rescue plan. This was obviously a significant event, but one that was neither sudden nor unexpected by the time it arrived. The initial turbulence of the Greek debt crisis had passed by this point, with Greece having accepted a bailout on 15 April 2010, and the next bailout crisis (in Hungary) was not to occur until July.

UK Hung Parliament – May 2010

[23] Source: http://www.theguardian.com/world/2013/jun/09/francois-holland-eurozone-crisis-over

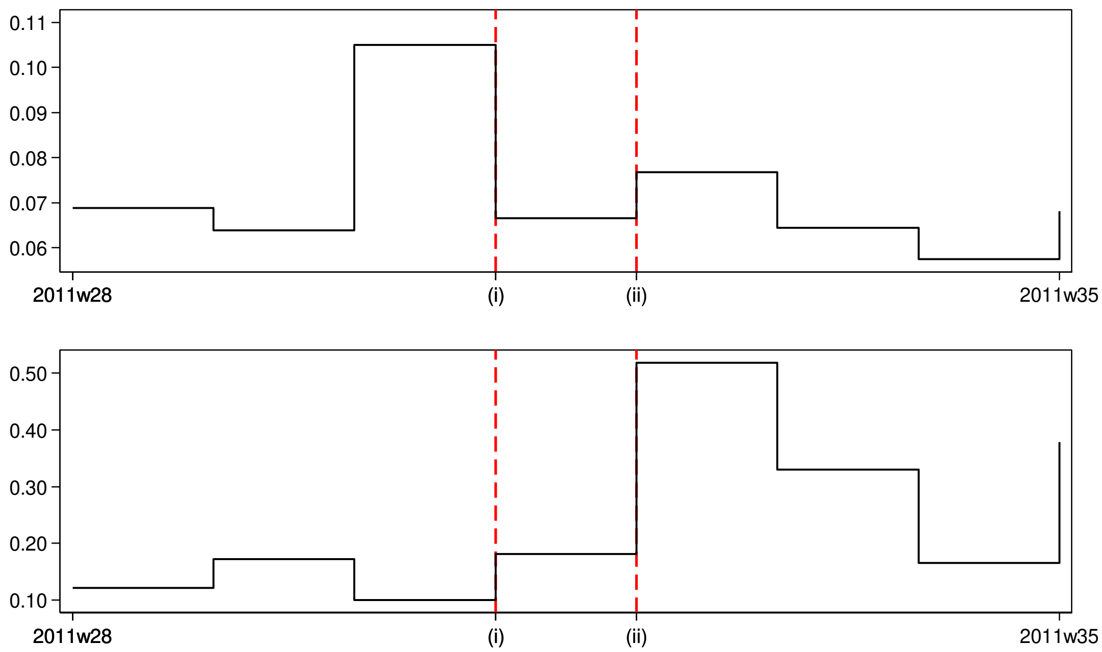
Weekly



(i) UK national elections result in hung parliament for first time since 1974 [6 May 2010]
(ii) Coalition government formed [12 May 2010]

The second sustained ramp up in U during the EZ crisis, from June to October 2011, coincides with events on both sides of the Atlantic. In the US, Congress was threatening to not extend the US government debt ceiling by the deadline of 2 August 2011, and thus force the US into default, generating uncertainty about the knock on effects for the global financial system. In the EZ, the government debt crisis that had already prompted bailouts for Greece, Ireland and Portugal was threatening to engulf Spain and Italy too. Thus on 8 August 2011 the ECB re-activated its Securities Markets Program to buy Spanish and Italian government bonds in large quantities. The intra-month timing in U suggests that the FT focussed more on uncertainty associated with the US debt ceiling deadline. σ does not indicate substantial uncertainty in advance of either event. Furthermore, σ declines in subsequent months, though remaining elevated.

US debt ceiling crisis and ECB action to stem contagion, August 2011



(i) US government debt ceiling deadline (2 August)
(ii) ECB starts buying Spanish and Italian bonds (8 August)

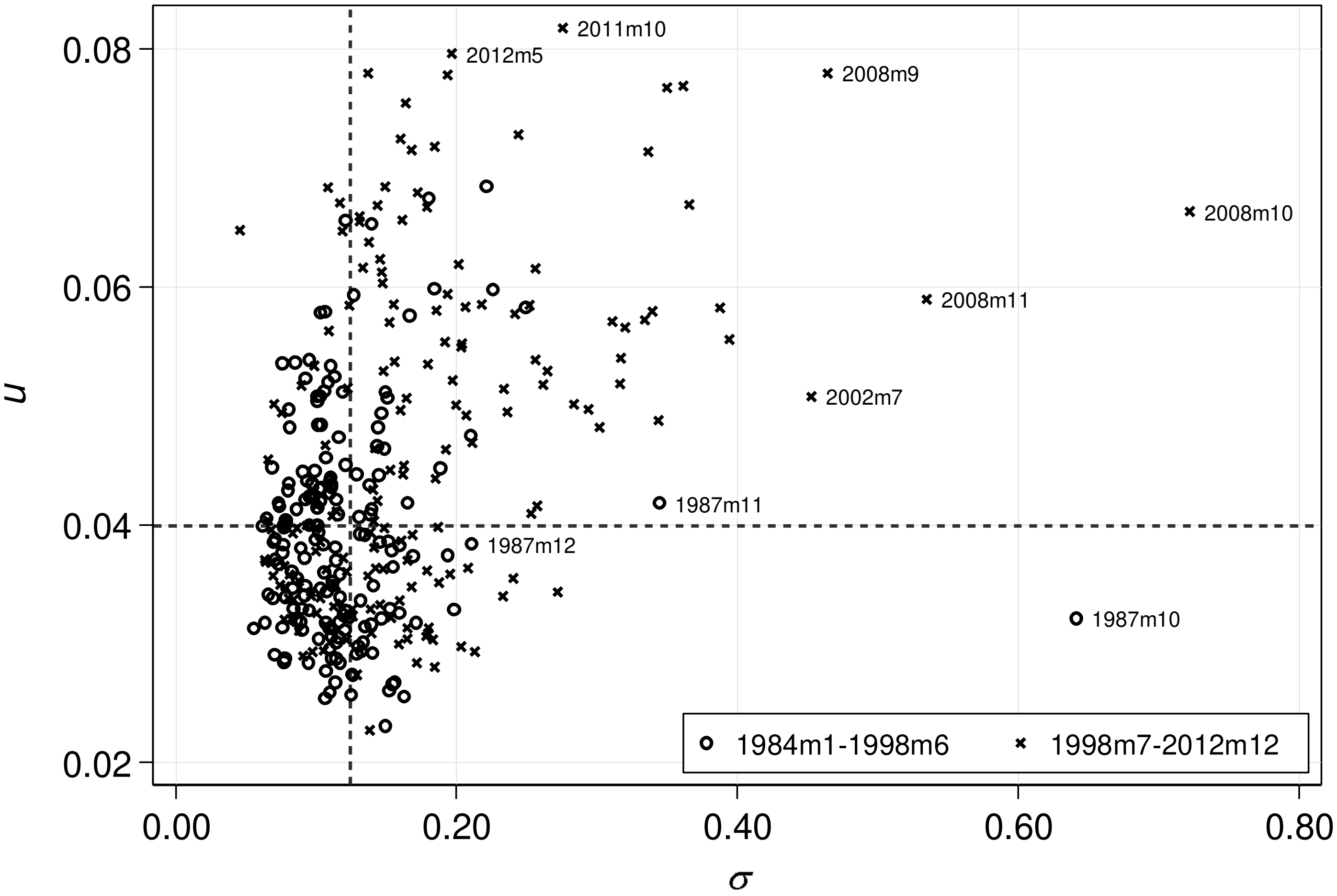
Finally, the jump up in U in May 2012, to its second highest value in the sample, coincides with a resurgence of the EZ crisis. On 6 May the Greek elections resulted in a majority for parties opposing the international bailout. However, they failed to form a coalition government, so new polls were announced for 17 June, leaving substantial uncertainty about the Greek bailout and therein the future of the Eurozone. On top of this, Spain's 4th largest bank asked for a bailout on 25 May.

6.3 Contemporaneous correlations

Cross-plots of U vs. σ and ΔU vs. $\Delta\sigma$ are shown in Figure 14 and Figure 15 below.

We do not show best-fit lines between U and σ for two reasons. First, standard regression methods assume that the independent variable is fixed and that the dependent variable is responsible for all variance around the best fit line, whereas no such asymmetry suggests itself in our context. Indeed, the gradient of best fit lines (including robust and non-parametric variants such as the Theil-Sen median slope, and the median regression line) varied widely depending of which of U and σ was cast as the dependent variable. Second, the magnitudes of regression slope parameters on these variables have no direct economic interpretation given the arbitrary nature of the scale on U.

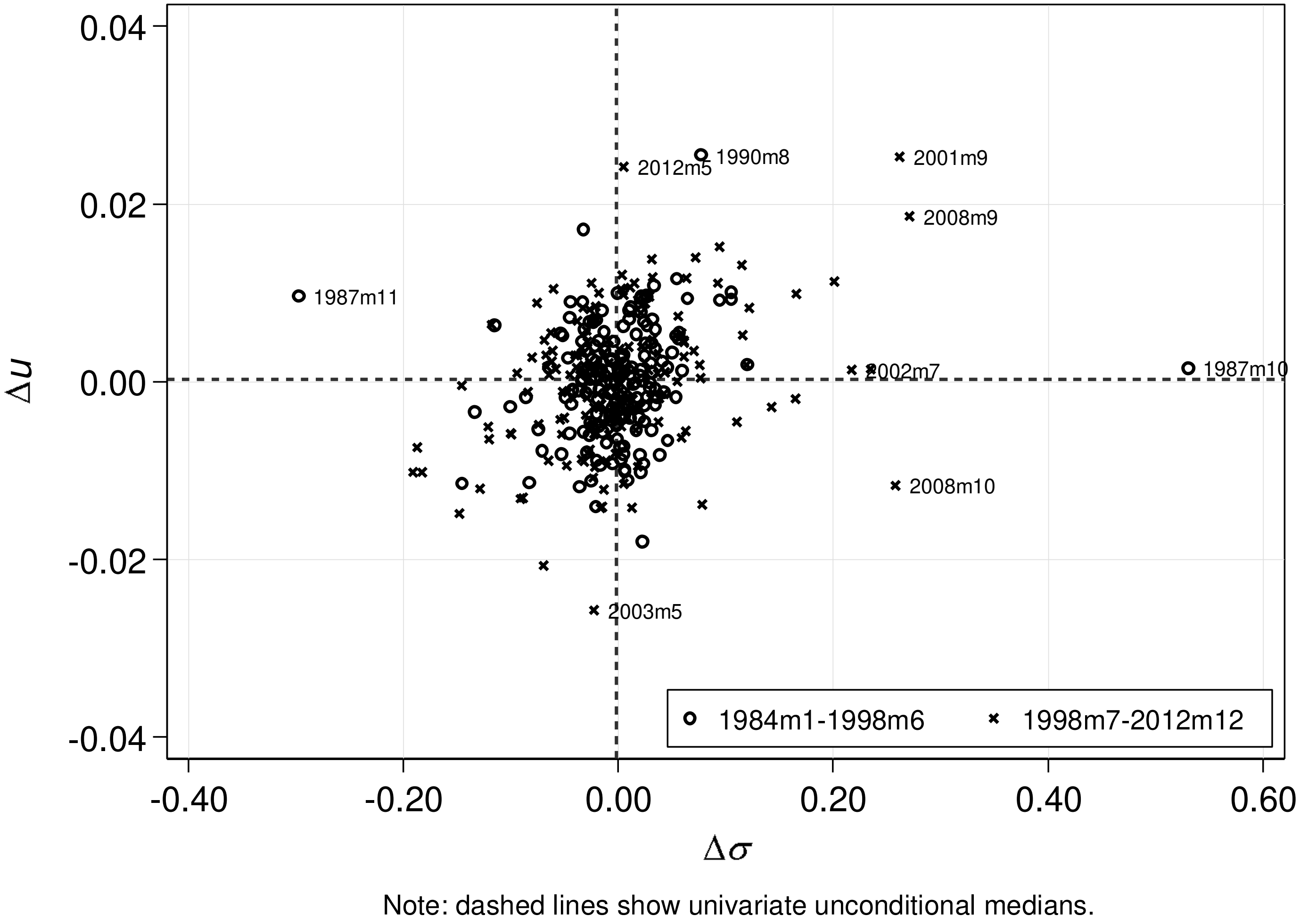
Figure 14: Cross-plot of levels of U and σ, 1984-2012



Note: dashed lines show univariate unconditional medians.

Notes: Observations from the first and second halves of the sample are given different marker symbols, to provide one rough visualisation of temporal change in the bivariate distribution. More of the largest observations fall in the second half of the sample period, and mostly relate to the financial crisis period at the end of the sample, and to the October 1987 crash.

Figure 15: Cross-plot of first differences ΔU and $\Delta\sigma$, 1984-2012



Notes: see notes to Figure 14.

6.3.1 Methodology

6.3.1.1 Measures of correlation

We use four measures of correlation/association – one cardinal, three ordinal – which make different trade-offs between efficiency and robustness.

Pearson's product-moment correlation ρ measures the degree of linear association between the variables. However, a priori we only have reason to expect the relationship to be monotonic, not necessarily linear. Also, ρ is sensitive to outliers, as seen, for example, in the abrupt jumps in rolling values of ρ in Figure 16.

Rank correlations are more outlier-resistant and capture the strength of monotonic association. Spearman's rank correlation r_s is a straightforward application of Pearson's ρ on ranks. Kendall's τ_a is calculated by permuting over all pairs of observations. It is the difference between the number of pairs whose ordering is the same according to both variables (concordance) and the number of pairs whose ordering is different (discordance), as a fraction of all pairs. r_s is perhaps more familiar via analogy to ρ, but more reliable methods for obtaining confidence intervals are available for τ_a than for r_s[24], and τ_a is easier to interpret precisely than r_s: τ_a equals the difference in probability of the two covariates 'agreeing' versus 'disagreeing', in the sense that for a randomly selected pair of observations the ordering of the observations (according which is larger) is the same according to both covariates.

Greiner's $r_g \equiv \sin\left(\frac{\pi}{2}\tau_a\right)$ transforms τ_a so as to render it comparable with ρ[25] while retaining the outlier-resistance, and invariance to monotonic transforms of U and σ, enjoyed by τ_a. If there exists a pair of monotonic transformations under which U and σ become bivariate normal (and we know that the log transform is already a reasonable approximation), then r_g is equal to the ρ that would be obtained on the transformed covariates, but without us having to know what those transformations are.

Sign concordance measures are still more robust to outliers, and to possible disturbances from non-uncertainty components of the measures, albeit at an efficiency cost. In the present context these are only useful for differences. For simplicity of interpretation we focus on the fraction of observations, $S \in [0,1]$, that have $sign(\Delta U) = sign(\Delta\sigma)$ where $sign(\cdot) \equiv \begin{cases} 1\ if & \cdot > 0 \\ 0\ if & \cdot \le 0 \end{cases}$.

6.3.1.2 Bootstrapped p-values and confidence intervals

HAC-robust variance estimators are not generally available for correlation statistics, and the 'classical' approximations for p-values and confidence intervals (CIs) assume independent observations and, in some cases, normally distributed data, which is clearly not congruent with the substantial and significant autocorrelation (see Table 4 and Figure 7) and non-normality (see Table 2) in our time-series data.

[24] This will matter more when estimating confidence intervals/p-values for the differences between degree of association in different subsets of the data, as is done below.

[25] Originally the one-to-one mapping between r_g and τ_a was derived under the assumption that the correlates are bivariate normal. However, as Newson (2002) highlights, it is not affected by odd-numbered moments (such as skewness) and is expected to hold approximately for a wide range of continuous bivariate distributions.

We therefore use the block bootstrap of Kunsch (1989), which resamples over non-overlapping blocks thus allowing for serial dependence within blocks. Bootstrap p-values are derived from bootstrap standard errors under the approximation that the sampling distribution is normal. Bootstrap CIs are the bias-corrected and accelerated CIs of Efron (1987) which do not assume a normal sampling distribution. All bootstrap results are based on 999 replications[26] and implemented using the -bootstrap- command in Stata.

Block length is important. For analysis in levels, where rolling autocorrelations are mostly damped to insignificance by 6 months, but sometimes persist longer, we use blocks spanning 6 months and 12 months (as a sensitivity check). This allows the majority of the serial dependence patterns to be captured within blocks (in light of the autocorrelation u and σ reported below) while also providing for a reasonable number of resampling units (key to an effective bootstrap). For first differences where autocorrelation is weaker, we use blocks spanning 2 month and 6 months.

Since the bootstrap is only asymptotically correct, we also show 'classical' p-values and CIs where they are available and space permits.

6.3.2 Full sample correlations

Correlation measures for the full sample are reported in Table 6, and sign concordances are reported in Table 7. In short, U and σ are strongly positively correlated.

Correlation in levels is positive, substantial (ρ on the order of 0.3-0.4), comparably sized at weekly through annual frequencies (daily results are discussed separately below), and statistically significant at the 99% level except for annual data where the confidence interval is wide due to the small number of observations. This is qualitatively similar to Baker et al.'s (2013) finding of a Pearson correlation of 0.578 between their economic policy uncertainty measure and the VIX for 1990-2012. Rank correlation coefficients are only slightly smaller than Pearson's ρ, indicating that ρ is not just driven by outliers. The CI of [0.087,0.304] on Kendall's τ_a for monthly data tells us that, with 95% confidence, for a randomly selected pair of months, U and σ are between 8.7% and 30.4% more likely to agree than to disagree in their ordering of the months.

Correlation in first differences is similarly positive, substantial, significant, robust, and comparably sized at monthly through annual frequencies (weekly results are discussed separately below). For a randomly selected pair of months, U and σ are between 9.7% and 24.6% more likely to agree than to disagree in their ordering. The sign concordances of first differences show a similar pattern across frequencies (first column of Table 7). U and σ move in the same direction in 57.1% of months, and this is significantly different, at the 99% level, to the 'coin-toss' benchmark of 50%. The probability of sign concordance does not vary strongly or significantly with the direction of movement (last four columns of Table 7).

The much weaker correlation at daily frequency, and in first differences at weekly frequency, is at least partly due to noisier estimates of σ at higher frequency, but *a priori* could also be partly due to dynamics causing interdependence to operate at lags and leads at higher frequency and thus not be reflected into contemporaneous correlations.

[26] This choice, rather than a round number such as 1000, avoids the need for interpolation in estimating results for conventional confidence levels.

We focus on monthly data for the deeper analysis of contemporaneous correlations in the remainder of this Section, and consider the higher frequency dynamics in Section 6.4.

Table 6: Contemporaneous cross-correlations in levels and first differences, 1984–2012

		N	Pearson's ρ		Spearman's r_s		Greiner's r_g		Kendall's τ_a	
Levels	daily	7326	0.126*** †††	*[0.103,0.148]* [0.062,0.204]^{6} [0.064,0.194]^{12}	0.089*** †††	*[0.066,0.112]* [0.041,0.151]^{6} [0.040,0.156]^{12}	0.093*** †††	*[0.069,0.117]* [0.042,0.158]^{6} [0.042,0.163]^{12}	0.059*** †††	*[0.044,0.075]* [0.027,0.101]^{6} [0.027,0.104]^{12}
	weekly	1508	0.305*** †††	*[0.258,0.350]* [0.167,0.419]^{6} [0.163,0.442]^{12}	0.237*** †††	*[0.188,0.283]* [0.123,0.344]^{6} [0.116,0.372]^{12}	0.245*** †††	*[0.194,0.294]* [0.129,0.356]^{6} [0.120,0.382]^{12}	0.157*** †††	*[0.124,0.190]* [0.082,0.232]^{6} [0.076,0.250]^{12}
	monthly	348	0.393*** †††	*[0.298,0.477]* [0.202,0.517]^{6} [0.188,0.544]^{12}	0.314*** †††	*[0.215,0.404]* [0.135,0.458]^{6} [0.105,0.499]^{12}	0.313*** †††	*[0.216,0.405]* [0.136,0.459]^{6} [0.103,0.497]^{12}	0.203*** †††	*[0.139,0.266]* [0.087,0.304]^{6} [0.066,0.331]^{12}
	quarterly	116	0.396*** †††	*[0.223,0.534]* [0.192,0.544]^{6} [0.161,0.565]^{12}	0.364*** †††	*[0.189,0.508]* [0.153,0.527]^{6} [0.135,0.556]^{12}	0.365*** †††	*[0.208,0.504]* [0.167,0.526]^{6} [0.136,0.555]^{12}	0.238*** †††	*[0.134,0.337]* [0.107,0.353]^{6} [0.087,0.374]^{12}
	annual	29	0.302 †	*[-0.094,0.588]* [-0.094,0.568]^{12}	0.249	*[-0.146.0.552]* [-0.159,0.552]^{12}	0.253	*[-0.128,0.568]* [-0.166,0.561]^{12}	0.163	*[-0.080,0.387]* [-0.106,0.377]^{12}
First differences	daily	7325	-0.025**	*[-0.047,-0.002]* [-0.056,0.008]^{2} [-0.055,0.014]^{6}	-0.020*	*[-0.043,0.003]* [-0.046,0.010]^{2} [-0.048,0.011]^{6}	-0.021*	*[-0.046,0.003]* [-0.049,0.010]^{2} [-0.050,0.012]^{6}	-0.014*	*[-0.029,0.002]* [-0.031,0.006]^{2} [-0.032,0.008]^{6}
	weekly	1507	0.049*	*[-0.001,0.099]* [-0.018,0.131]^{2} [-0.023,0.126]^{6}	0.028	*[-0.022,0.078]* [-0.044,0.087]^{2} [-0.044,0.089]^{6}	0.030	*[-0.026,0.086]* [-0.047,0.093]^{2} [-0.047,0.096]^{6}	0.019	*[-0.017,0.055]* [-0.030,0.059]^{2} [-0.030,0.061]^{6}
	monthly	347	0.272*** †††	*[0.171,0.366]* [0.159,0.394]^{2} [0.119,0.408]^{6}	0.243*** †††	*[0.140,0.338]* [0.144,0.356]^{2} [0.113,0.378]^{6}	0.256*** †††	*[0.142,0.365]* [0.151,0.377]^{2} [0.118,0.401]^{6}	0.165*** †††	*[0.091,0.238]* [0.097,0.246]^{2} [0.076,0.263]^{6}
	quarterly	115	0.414*** †††	*[0.243,0.549]* [0.239,0.555]^{3} [0.297,0.518]^{6}	0.418*** †††	*[0.248,0.552]* [0.237,0.570]^{3} [0.258,0.558]^{6}	0.447*** †††	*[0.254,0.605]* [0.257,0.603]^{3} [0.277,0.584]^{6}	0.295*** †††	*[0.164,0.415]* [0.164,0.412]^{3} [0.178,0.397]^{6}
	annual	28	0.283	*[-0.113,0.575]* [-0.082,0.680]^{12}	0.220	*[-0.175.0.531]* [-0.150,0.585]^{12}	0.245	*[-0.198,0.605]* [-0.143,0.638]^{12}	0.158	*[-0.126,0.417]* [-0.091,0.441]^{12}

Notes: N is the number of observations. *, **, *** indicate significance at 10%, 5% and 1% levels respectively, based on approximate two-sided 'classical' p-values (see below). Similarly †, ††, ††† indicate significance but based on normal block bootstrap two-sided p-values (with block size the smaller of those used in constructing CIs) using the methods described in Section 6.3.1.2. Brackets [] contain 95% confidence intervals (CIs). 'Classical' CIs are *italicised*. Block bootstrap CIs are non-italicised and the block length in months is noted in a superscript suffix. Shorter blocks are used for first differences than for levels because autocorrelation is shorter-lived in first differences. 'Classical' p-values and CIs for ρ and r_s are obtained using Fisher's approximation, which assumes that observations are independent and bivariate normally distributed and the sample size is large. Under these assumptions, the sampling distributions of the hyperbolic arctangents of ρ and r_s (Fisher's z-transform) are asymptotically normal with mean $\rho + 2\rho/(N-1)$ or $r_s + 2r_s/(N-1)$ respectively (we subtract the second term in each expression as a bias adjustment) and variance approximately equal to $1/(N-3)$. 'Classical' p-values and CIs for τ_a and $r_g \equiv \sin(\pi\tau_a/2)$ are obtained by the method of Newson (2005), using the jackknife and Taylor polynomial approximations, after applying Fisher's z transform to stabilise variances as recommended in Edwardes (1995).

Table 7: Contemporaneous sign concordance in first differences, 1984–2012

	N	All		$\Delta U \leq 0$		$\Delta U > 0$		$\Delta\sigma \leq 0$		$\Delta\sigma > 0$	
daily	7325	0.498	$[0.485,0.512]^{2}$ $[0.485,0.511]^{6}$	0.498	$[0.485,0.512]^{2}$ $[0.485,0.511]^{6}$	0.498	$[0.483,0.514]^{2}$ $[0.483,0.513]^{6}$	0.503	$[0.486,0.517]^{2}$ $[0.489,0.517]^{6}$	0.493	$[0.479,0.508]^{2}$ $[0.479,0.507]^{6}$
weekly	1507	0.518	$[0.489,0.546]^{2}$ $[0.486,0.544]^{6}$	0.515	$[0.481,0.546]^{2}$ $[0.476,0.542]^{6}$	0.520	$[0.488,0.552]^{2}$ $[0.491,0.557]^{6}$	0.515	$[0.480,0.547]^{2}$ $[0.481,0.546]^{6}$	0.520	$[0.485,0.553]^{2}$ $[0.487,0.553]^{6}$
monthly	347	0.571†††	$[0.523,0.624]^{2}$ $[0.513,0.631]^{6}$	0.598 †††	$[0.520,0.671]^{2}$ $[0.524,0.667]^{6}$	0.545	$[0.476,0.612]^{2}$ $[0.472,0.616]^{6}$	0.555†	$[0.483,0.621]^{2}$ $[0.481,0.619]^{6}$	0.588††	$[0.512,0.653]^{2}$ $[0.519,0.656]^{6}$
quarterly	115	0.635†††	$[0.530,0.722]^{3}$ $[0.526,0.713]^{6}$	0.679 †††	$[0.553,0.789]^{3}$ $[0.536,0.796]^{6}$	0.593	$[0.462,0.719]^{3}$ $[0.474,0.704]^{6}$	0.613†	$[0.434,0.726]^{3}$ $[0.471,0.721]^{6}$	0.660††	$[0.509,0.776]^{3}$ $[0.526,0.780]^{6}$
annual	28	0.607	$[0.393,0.750]^{12}$	0.583	$[0.308,0.875]^{12}$	0.625	$[0.375,0.857]^{12}$	0.538	$[0.231,0.778]^{12}$	0.667	$[0.353,0.857]^{12}$

Notes: N is the number of observations at the given frequency. In each of the remaining columns the left-most number is the fraction S of observations for which $sign(\Delta U) = sign(\Delta\sigma)$, with $sign(\cdot)$ as defined in the main text. †, ††, ††† indicate significant difference from 0.5 at 10%, 5% and 1% levels respectively, based on block bootstrap p-values (described in Section 6.3.1.2) for the shorter of the block lengths used for the CIs. Brackets [] contain block bootstrap 95% confidence intervals (CIs) and the block length in months is noted in a superscript suffix. CIs are shown for different block lengths to allow assessment of their sensitivity (or lack thereof) to block length.

6.3.3 Temporal variation in correlations

The full sample correlations conceal substantial temporal variation, as clearly illustrated in Figure 16 using rolling 60-month windows[27]. Occasional large jumps in Pearson's ρ betray its sensitivity to outliers, but the more robust rank correlation measures exhibit similar temporal profiles to ρ (albeit with less abrupt changes), as do correlations in first differences and sign concordance of first differences.

The correlation between U and σ drops off sharply from early 2010, falling to nearly zero by the end of the sample at end 2012. By contrast, the correlation and sign concordance between ΔU and $\Delta\sigma$ rose through 2010 and remained high to the end of the sample. The timing coincides with the start of the Eurozone crisis, conventionally dated to the emergence of the Greek government debt crisis in late 2009/early 2010.

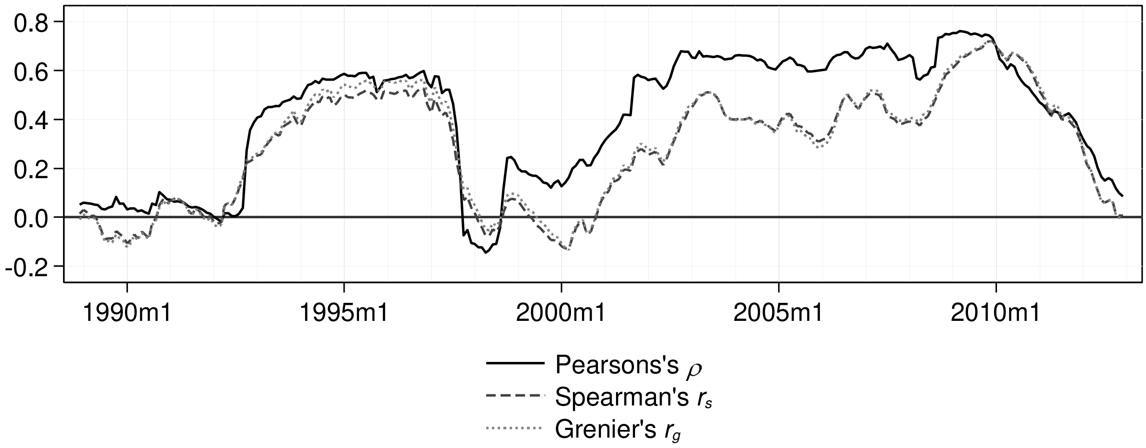
One potential explanation might be that the extraordinary interventions by fiscal and monetary authorities during the Eurozone crisis artificially supressed the level of σ (even while movements around the supressed base level continued to reflect movements in uncertainty), whereas the news-media continued to express the strongly elevated level of uncertainty, causing a disconnect in levels between U and σ. Of course this is speculative, and does not account for the fact that the decline in σ began earlier in 2009, nor why earlier interventions such as TARP in the US, and the £200billion of asset purchases undertaken by the UK authorities between March 2009 and January 2010, were not associated with a similar decline in correlation.

The correlation between U and σ also exhibits switching behaviour between sustained periods of very low correlation and sustained periods of very high correlation. The consistency of this pattern across rank and sign concordance measures demonstrates that this is not an artefact of outliers moving in and out of the rolling window. The sustained nature of the movements, even allowing for the smoothing inherent in a rolling analysis, suggests some underlying structure.

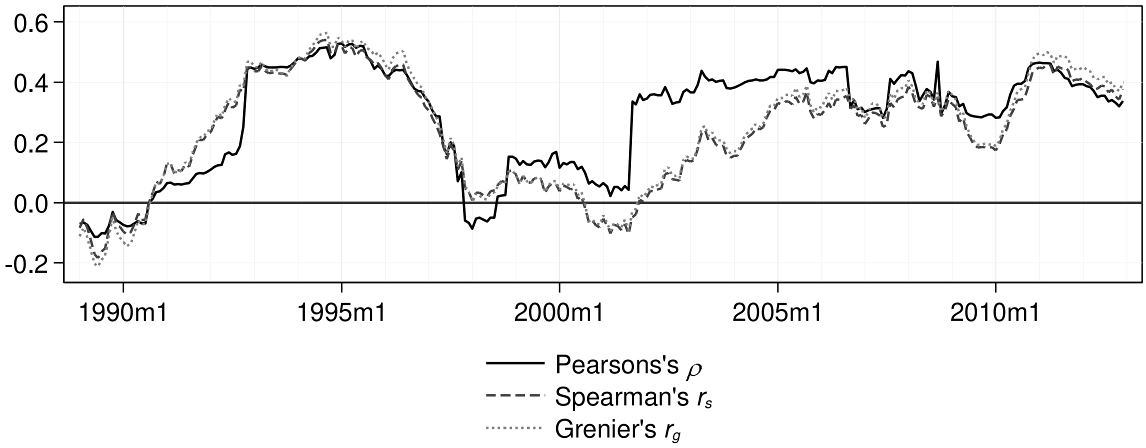
[27] This is the same window lengths used by Campbell et al. (2001) in their investigation of the relationship between disaggregated components of volatility. Five-year rolling correlations on weekly and quarterly data show similar time profiles, but the rise in correlation of levels around the onset of the financial crisis and the subsequent drop is less pronounced (relative to the early-mid 1990s hump) at lower frequency.

Figure 16: Contemporaneous cross-correlations within rolling 60-month windows

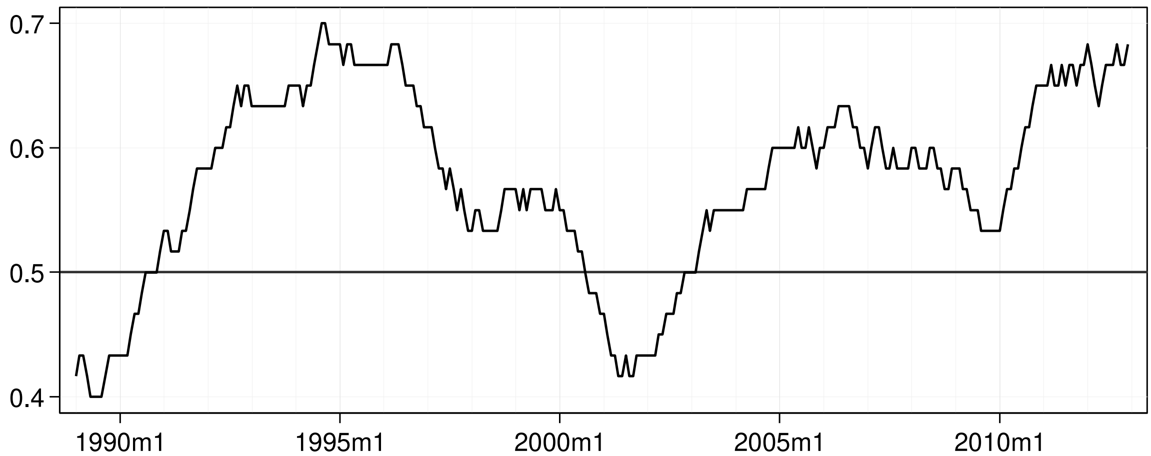
Levels, U and σ



First differences, Δu and $\Delta\sigma$



Fraction of observations, S, with sign agreement between ΔU and $\Delta\sigma$



Notes:

1. date axes indicate end date of the five-year rolling window.
2. τ_a and r_g are related to one another by a time-invariant monotonic transform, so there is no incremental information from displaying both. We display r_g because it is more intuitively comparable with ρ and r_s.

What might cause this switching behaviour? We offer two hypotheses, which are not mutually exclusive.

First, journalists may have a lower propensity to use the word “uncertainty” to express upside risk than to express equally sized downside risk. U would then tend to underweight upside risk whereas σ would weight upside and downside risk equally. Correlation between U and σ may then be weaker in periods when upside risk is higher. Since perceived upside risk is probably higher during booms, we would expect lower correlation during booms. This is broadly consistent with what we observe: the periods of lower correlation broadly correspond respectively to the late 1980s macroeconomic upswing known as the ‘Lawson boom’ (after the then UK Chancellor) and to the dot-com boom of the late 1990s.

A few recent papers that attempt decompose uncertainty into upside and downside components, provide some corroborating evidence for our hypothesis, albeit with different empirical uncertainty measures than used here. Rossi & Sekhposyan (2015) identify uncertainty with a measure of the size of realised error[28] on GDP forecasts, which the authors interpret as a measure of unpredictability, which they further assert is associated with uncertainty. Two of the three extended periods of upside uncertainty shown in their results for our sample period coincide broadly with the periods of low correlation identified above. The third, around 1992-1993 does not. (See their Figure 2, second and fourth panels.) That said, they use US rather than UK data. Segal, Shaliastovich, & Yaron (2014) identify ‘good’ (‘bad’) or upside (downside) uncertainty with positive (negative) realised semi-volatilities of the US industrial production growth rate, using the estimator of Barndorff-Nielsen, Kinnebrock, & Shephard (2008). They also show upside uncertainty being elevated in the mid to late 1990s, at to a lesser extent in the late 1980s, relative to the rest of our sample period (see their Figure 2). Feunou, Jahan-Par, & Tedongap (2010) show similar timings for elevated S&P500 upside volatility, estimated from a binormal-GARCH model (see their Figure 2, Panels E and F).

Future research could test our hypothesis more rigorously by applying the present correlation framework to the semi-volatilities of aggregate stock index returns, and through manual semantic analysis of occurrences of “uncertain*” in FT articles published in the late 1990s about dot-com stocks[29].

Second, risk aversion (distinct from the perceived degree of risk itself) has been estimated to account for perhaps one quarter of stock returns volatility (Bekaert, Engstrom, & Xing, 2009). Temporal variation in risk aversion might therefore contribute to temporal variation in the correlation between U and σ. Future research could use the framework of Bekaert, Engstrom, & Xing (2009) to decompose σ into risk aversion and risk components, and examine how these relate to U.

[28] Specifically, the difference between the 0.5 (representing the median) and the quantile of the realised forecast error with respect to its historic distribution

[29] An outline of such a research program might be: i) manually tag statements that express or imply uncertainty about the future; ii) manually classify these by direction of risk (upside/downside/symmetric) and index the words or phrases used; iii) examine correlations between direction of risk and particular words or phrases.

6.3.4 Further structure in the correlation

Finally, we document how the correlations between the movements of U and σ, and their signs and magnitudes, vary with the level of U and σ, and with the magnitude and direction of their movements. The aim is to establish basic empirical facts against which future theories of the relationship between U and σ can be developed and tested.

We segment observations into a series of two-by-two grids[30] and estimate correlation and its significance within each cell and in the corresponding margins, along with the p-value for the difference between each pair of cells or margins. Figure 17 and Figure 19 segment the sample by whether the lagged levels $L.U$ and $L.\sigma$ are above (or at) or below their medians (referred to as 'high' and 'low' for ease of exposition). We use lags because contemporaneous levels are by construction correlated with the first differences, the variation in whose correlation we are trying to measure. Figure 18 segments by whether the magnitude of first differences $|\Delta U|$ and $|\Delta\sigma|$ are above (or at) or below their medians. Figure 20 segments the sample by the sign of first differences.

Correlation in movements and concordance in their direction are substantial and significant when $L.U$ is high (bottom rows of Figure 17 and Figure 19) but weak and insignificant when the level $L.U$ is low (top rows). This true whether $L.\sigma$ is high or low, and thus also in the margin[31]. By contrast, these measures of co-movement do *not* vary strongly or significantly when depending on whether $L.\sigma$ is high versus low.

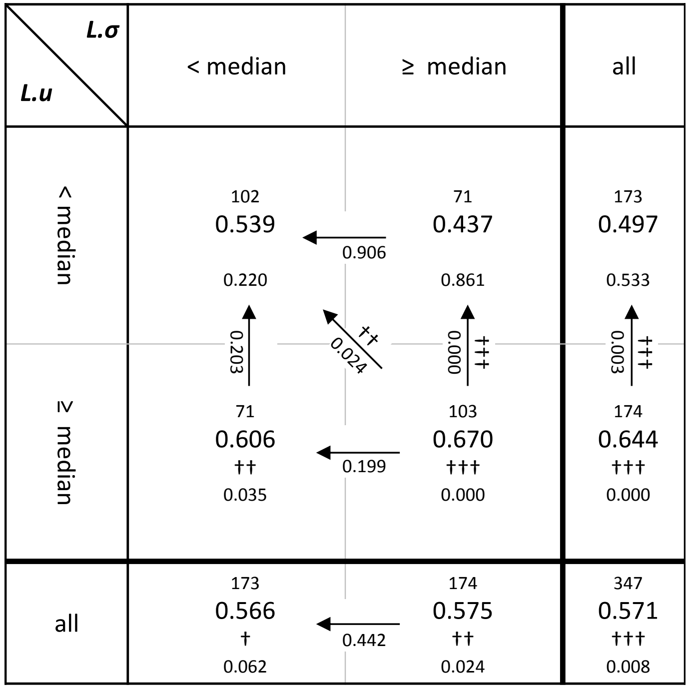
The pattern is similar for concordance in the direction of movements when segmenting by the magnitude of movements $|\Delta U|$ and $|\Delta\sigma|$. There is significantly stronger concordance when there is a large movement U compared to essentially no concordance when the movement in U is small. By contrast, the magnitude of movement in σ does not make a significant difference.

The lack of concordance when *both* U and σ are low might be partly due to a lower signal-to-noise ratio, but the asymmetry in dependence of correlation/concordance on U vs. σ demands a more structural explanation.

The magnitude of movements in U and σ is fairly well correlated when they are moving in the same direction (diagonal cells in Figure 20) – more strongly so when they are both falling rather than rising, though the difference is not significant. Unsurprisingly, correlation in magnitude of movements is weaker when the measures move in opposite directions.

[30] Experiments with more granular segmentation left too few observations per cell to conduct useful inference.
[31] In Figure 19, the value of $\rho = 0.278$ when $L.U$ is low and $L.\sigma$ is high (and thus also the marginal $\rho = 0.144$ for $L.U$ low) is driven by outliers. The more robust rank correlation measures show much smaller values here.

Figure 17: Segmentation of sign concordance S, by lagged levels of U and σ



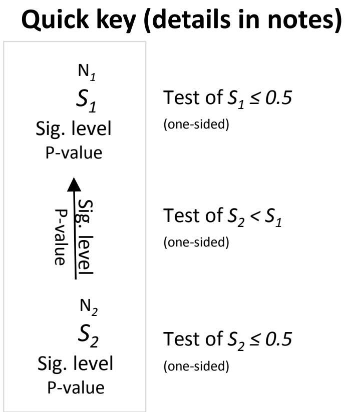
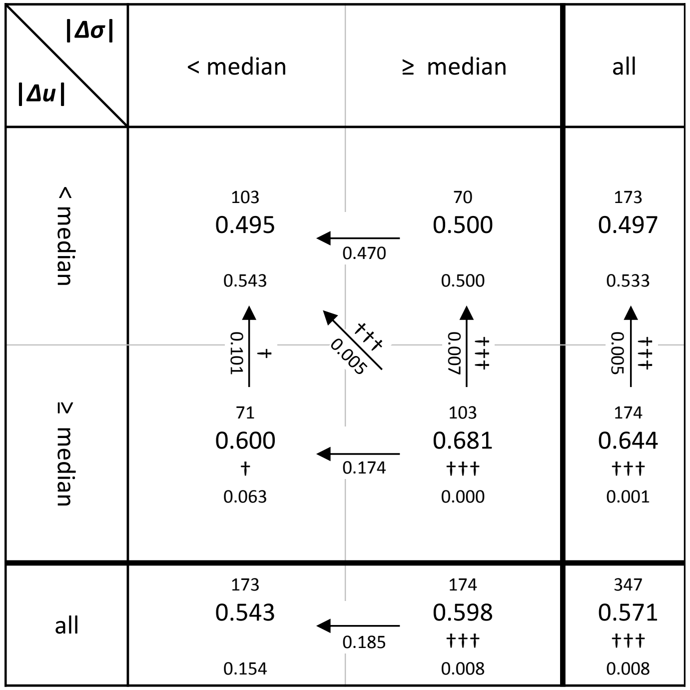


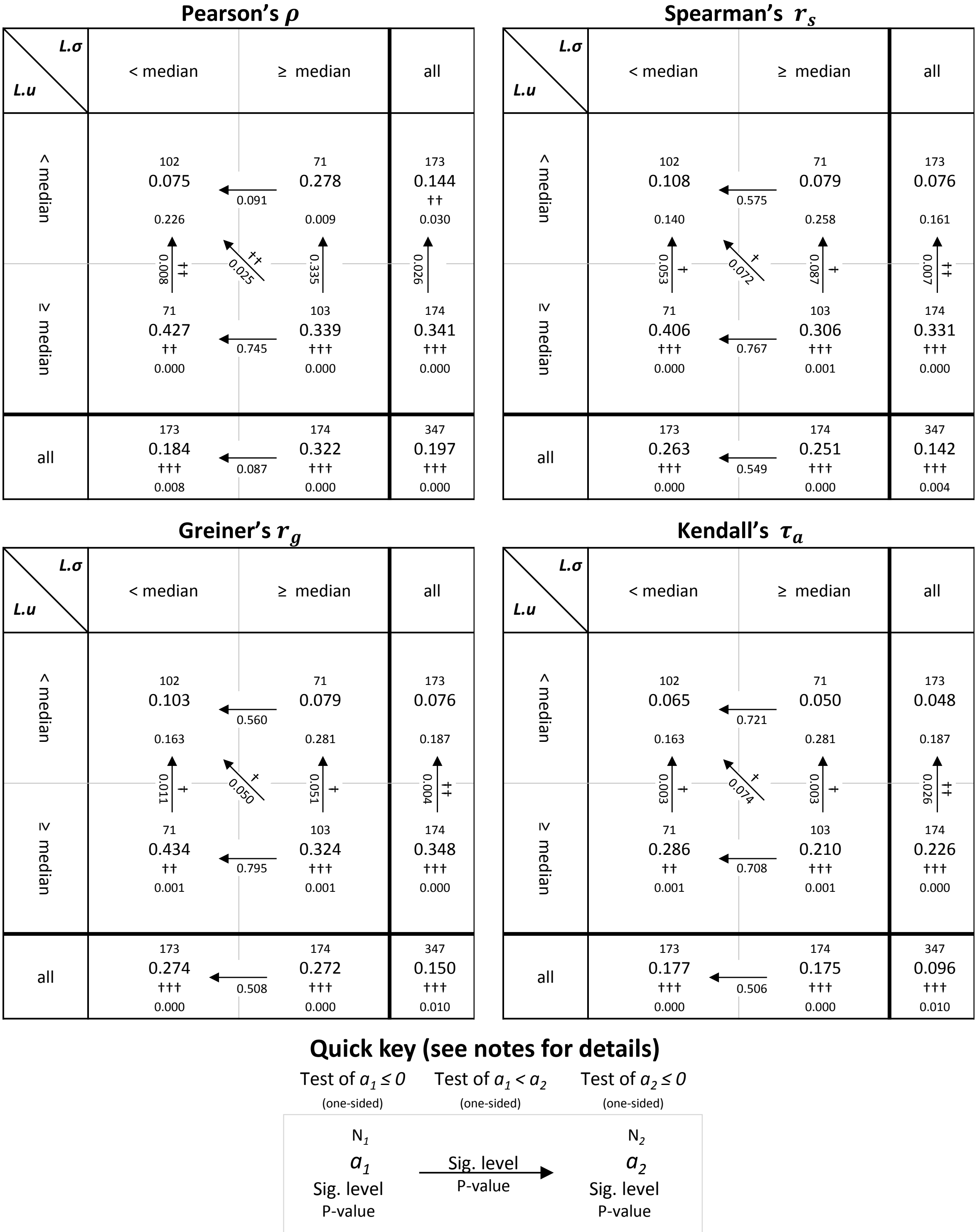
Figure 18: Segmentation of sign concordance S, by size of movements $|\Delta U|$ and $|\Delta\sigma|$



Notes: Within cells and margins: Reported p-values are one-sided for the null that the fraction of observations, S, exhibiting sign agreement is less than 0.5. These p-values are derived from the normal block bootstrap (see Section 6.3.1.2) using 6-month blocks. †, ††, ††† indicate significant rejection of the same null at 10%, 5% and 1% levels respectively, but based on whether 0.5 lies beyond the bias-corrected and accelerated block bootstrap confidence limit.

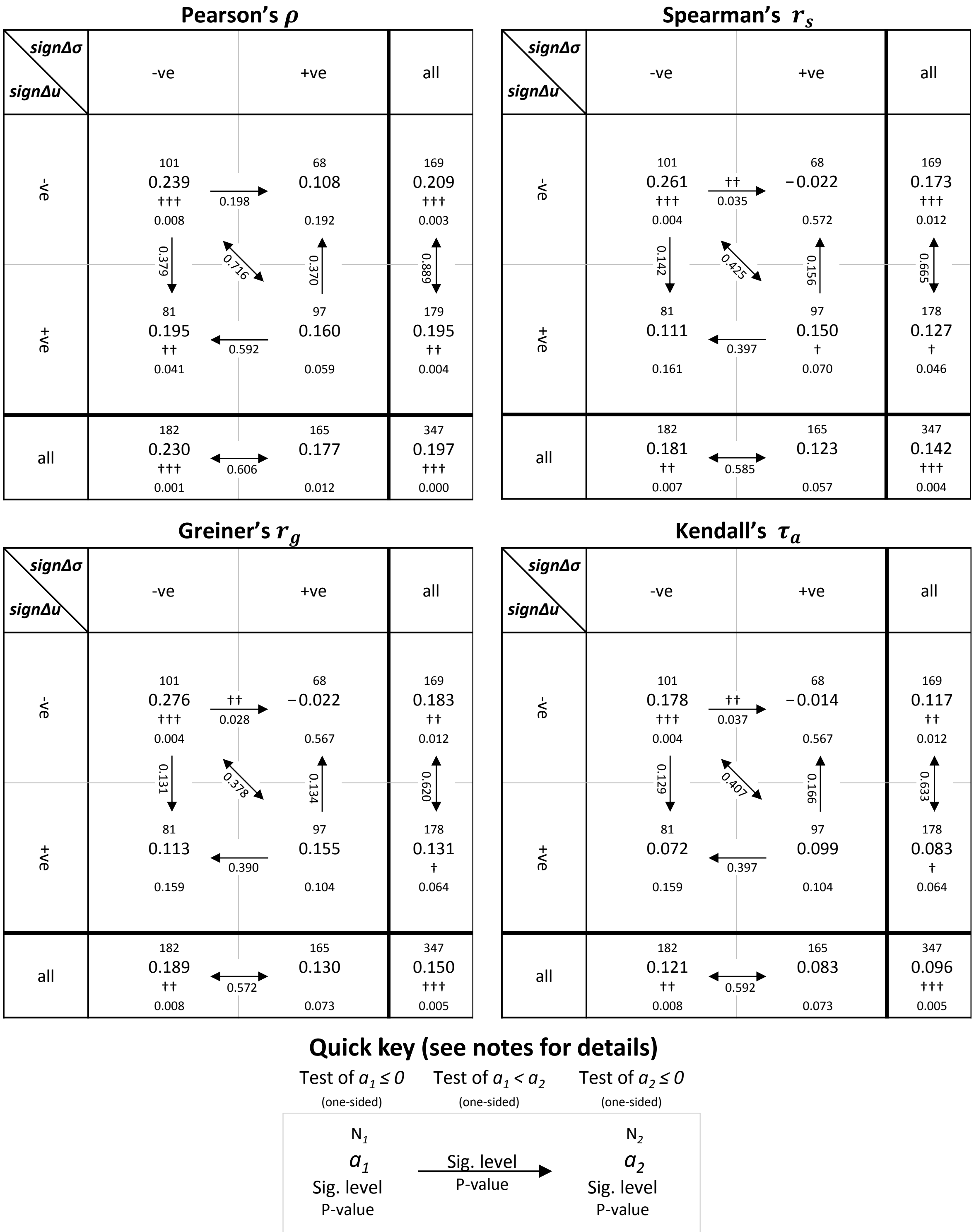
Comparisons between cells and between margins (indicated by arrows): Reported p-values are one-sided for the null that S, in the cell (or margin) at the arrow's tail, is less than in the cell (or margin) at the arrow's head. The choice of one-sided p-values reflects our prior expectation that S will be higher when the segmenting variables are larger. †, ††, ††† indicate significant rejection of the same null, at 10%, 5% and 1% levels respectively, but based on whether zero is contained within one-sided block bootstrap bias-corrected and accelerated confidence limits.

Figure 19: Correlation of ΔU and $\Delta\sigma$, segmented by level of U and σ



Notes: <u>Within cells and margins:</u> Reported p-values are one-sided for the null that the association measure is less than zero. These p-values are derived 'classical' method s described in notes to Table 6, assuming independence of observations. †, ††, ††† indicate significant rejection of the same null at 10%, 5% and 1% levels respectively, but based on whether zero is lies beyond the bias-corrected and accelerated block bootstrap confidence limit, which allows for serial dependence within 6-month blocks (see Section 6.3.1.2).
<u>Comparisons between cells and between margins (indicated by arrows):</u> Reported p-values are one-sided for the null that the measure of association, in the cell (or margin) at the arrow's tail, is greater than in the cell (or margin) at the arrow's head. The choice of one-sided p-values reflects our prior expectation that association will be higher when the segmenting variables are larger. †, ††, ††† indicate significant rejection of the same null at 10%, 5% and 1% levels respectively, but based on whether zero is contained within one-sided block bootstrap bias-corrected and accelerated confidence limits.
<u>Outliers</u> are behind the $\rho = 0.278$ value for $L.U <$ median and $L.\sigma \geq$ median, but do not so strongly affect the other more outlier-resistant correlation measures.

Figure 20: Correlation of $|\Delta U|$and $|\Delta\sigma|$, segmented by sign of movements $sign(\Delta U)$ and $sign(\Delta\sigma)$



Notes: see notes to Figure 19, with the exception that p-values and CIs for comparisons on the diagonal and in the margins are two-sided (indicated by a bidirectional arrow) because we lack a clear prior on the sign of the differences.

6.4 Granger causality tests

The correlated components of U and σ may be plausibly interpreted as components of latent uncertainty that are reflected in both measures. In this Section we ask whether this information is incorporated more quickly into one measure than into the other, or equivalently whether one or both of the measures can help to forecast the other.

To investigate this we test for Granger causality in a bivariate vector autoregression (VAR), in a similar spirit to Campbell et al. (2001) (albeit they compare aggregate stock volatility with industry- and firm-level volatility, rather than with news-media uncertainty). *A priori* we expect any Granger causation to be visible only at high frequency, given the strong incentives for rapid incorporation of information in both the news media and the stock market, so we conduct analysis at daily frequency, using realized volatility estimates of σ (see Section 6.1.2).

6.4.1 Baseline specification

To ensure a sample without gaps, and to simplify analysis, the calendar of 'days' is constructed to include only the 3 247 days that are both London Stock Exchange trading days and FT publication days, so that the 'day' after Friday is Monday (or Tuesday or Wednesday on Bank Holiday weekends) and we discard Saturday editions of the FT. Note that the distribution of U is little affected by whether Saturdays are included or not (see Table 2 on pg. 23) and the information loss is likely modest because if latent uncertainty is very high on a non-trading day then the strong autocorrelation documented above means it is also more likely to be high on the next trading day and thus expressed in U.

We apply skew-reducing transforms $\sqrt{U}$ and $\ln(\sigma)$ to render the variables approximately normal before entering them in the VAR. This mitigates the misspecification that would otherwise arise from the right skew in $U$and σ feeding through to the VAR residuals (which account for around 60% of the variance of U and 20% for σ).

We augmented the VAR with a full set of day-of-week dummies to control for weekly seasonality in U which, though an order of magnitude smaller than the interquartile range of the measure, was not trivial[32]. To these we added a small number of individual day dummies[33] for the largest and visually obvious outliers to mitigate possible distortion of inference. Estimates were very similar, and our conclusions unchanged, if these dummies were omitted.

Maximum lag length was set to 11 based on the Hannah-Quinn information criterion.

No major misspecifications are suggested by standard residual diagnostic plots (see Figure 21, Appendix C). No serious parameter instability is exhibited by the rolling parameter estimates in Figure 22. As anticipated by Granger (1998), some of our formal misspecification tests reject the null because we have orders of magnitude more observations and thus greater power than in the typical quarterly macroeconomic VAR. However, the *size* of misspecifications is modest. Residual serial correlation coefficients at the first lag are −0.008 and −0.006 for the U and σ equations respectively, and are similarly small at longer lags (see residual diagnostic plots). Residual skewness is −0.077 and 0.321, comparable to levels deemed acceptable in the leading reference text on VARs by Juselius (2006). Residual kurtosis is 3.24 and 3.91, not too far from the normal value of 3.

[32] This is seasonality that remains even after normalising for news volume as discussed in Section 5.4.
[33] 17/04/2000, 17/07/2000, 11/09/2000, 22/05/2001, 11/09/2001, 26/06/2002, 20/09/2002, 07/07/2005, 07/11/2007, 08/09/2008, 24/08/2011, 17/01/2012, 28/12/2012

6.4.2 Results and interpretation

We find no evidence of Granger causation from $U \to \sigma$: the p-value for a Wald test of the joint exclusion restriction on all lags of U in the σ equation is 0.487.

There is however statistically significant evidence of Granger causation from $\sigma \to U$: the p-value for a Wald test of the joint exclusion restriction on all lags of σ in the U equation is 0.010. This is primarily attributable to the first daily lag of σ in the U equation, for which the parameter is around twice as large as subsequent lags and has a p-value of 0.006. All subsequent lags are individually and jointly insignificant at the 95% level (joint p-value 0.202).

This tidy lag pattern is consistent with the intra-day timing of events: the FT is published in the morning before the markets open (recall our dataset corresponds to the print edition), thus uncertainty events occurring later in the day cannot be reflected in U until the following day, but such events occurring before market close may be reflected in σ that same day.

The joint dynamics are heavily dominated by autoregression. The statistically significant cross-equation effect identified above is very small in magnitude: the estimated response of U to a one standard deviation impulse in σ never exceeds 0.006 standard deviations of U.

It is conceivable that a richer econometric specification might capture some incremental forecasting power of U for σ, which would be of interest in risk management and quantitative trading strategies. However the lack of significant Granger causation $U \to \sigma$ in our linear VAR suggests that any such gains are likely to be modest.

Combined with the earlier evidence of strong correlation between the measures, these Granger causation results are consistent with the hypothesis that U and σ rapidly and completely incorporate the information from some common latent variable, which we suggest has natural interpretation as a composite of latent uncertainty components.

6.4.3 Robustness

These results, including the pattern of parameter estimates and conclusions from the Granger causality tests, were robust to alternative maximum lag lengths (including 5 and 22 as selected by Schwarz Information Criterion and the Akaike Information Criterion), stopping the sample before 2007m7 (an early conventional date for the start of the financial crisis) and indeed subsampling in many other ways as seen in the rolling analysis (Figure 22, Appendix C), omission of dummies, and including Saturday's articles in the calculation of U for the following Monday instead of dropping them.

7 Conclusions

We have provided a general framework for measuring latent uncertainty about a specified subject using news-media textual data, and a basic empirical implementation for aggregate uncertainty. We have also provided improvements to existing methodology, and demonstrated the importance of de-duplication and normalising uncertainty article counts by time-varying news volume.

Overall, our empirical analysis of U suggests that it is plausible proxy for aggregate uncertainty. It moves in ways that one would expect of latent uncertainty both in the broad sweep of the last thirty years, including around major narrative events that are conventionally associated with elevated uncertainty. U is also strongly, significantly and robustly correlated with the longer established uncertainty proxy, stock returns volatility. The Granger causation between them is largely attributable to intra-day timing (publication of the FT before the markets open), which is consistent with the hypothesis that U and σ track a common underlying latent variable, naturally interpreted as (a common component of) latent uncertainty, and that both measures incorporate the corresponding information efficiently. That said, the counter-intuitive decline in the level of σ after the on-set of the recent financial crisis illustrates how σ may be susceptible to artificial suppression due to invention by the official sector, so that U may provide a more reliable guide to the level of uncertainty at times of major financial dislocation.

That said, the switching behaviour between sustained periods of high correlation versus very low correlation, begs explanation. We have speculated on two potential causes – a bias in the semantics of "uncertain*" towards downside uncertainty, and time-varying risk-aversion – and outlined how these could be tested in future research. We have also documented further structure in the relationship between U and σ that can inform future model building and testing.

These results provide empirical foundations for the emerging literature that uses similar news-media uncertainty measures, and illuminate the path to further developing such measures. From the perspective of the literature on stock volatility, these results also support the thesis that volatility is connected to uncertainty about fundamentals (which are here captured in news-media references).

8 Directions for future research

The literature on news-media uncertainty measures is young and there are many areas in which advances could be sought.

The comparative analysis above could be extended. Deeper probing of the current data might focus on the switching behaviour apparent in the correlations between U and σ, using a Markov switching model, and the dependence of the correlation on the level of U, using a threshold VAR or smooth transition VAR. Developing convincing structural models is however likely to require a much richer empirical understanding of the cognitive economic models of journalists and market participants. The scope of the comparative analysis could be extended to include other extant uncertainty proxies such as those in Haddow et al. (2013), and other textual corpora, notably newswires which are more structured than blogs, and contain more incremental information relative to the FT than other mainstream media sources.

The measurement methodology could be refined in several areas, including the de-duplication algorithm and the standardisation of the definition of an 'article' over time, but we expect the greatest gains lie in developing the classifiers, and we end by discussing these.

8.1 Developing the uncertainty classifier

Basing the uncertainty classifier on simple keyphrase searches has yielded sensible results, and avoids the complexity associated with deeper parsing of the article text. We therefore suggest three next steps in this direction.

First, the semantics of "uncertain*" references in the FT should be investigated. We have hypothesised that "uncertain*" tends to be used more to express uncertainty to the downside than to the upside. As discussed in Section 6.3.3, this could be tested indirectly by comparing to directional semi-volatilities corresponding to σ, and directly by a manual semantic analysis, which might focus on dot-com stocks in the late 1990s where perceived upside risk was significant.

Second, additional uncertainty-related keyphrases should be explored. Conceivably, in addition to boosting signal strength, these might partially fill out gaps in the semantic range of "uncertain*". However, careful semantic analysis is even more important for less obvious keyphrases, to understand what noise is also being introduced.

Third, uncertainty classifiers could be designed to account for the relative semantic force of different keyphrases, their frequency and prominence of place within the article (e.g. headline vs. tail paragraphs), and qualifiers of degree (e.g. "very"). Such classifiers could be calibrated to a human scored sample articles using regression or supervised machine learning.

Our initial experimentation with additional uncertainty keyphrases may help direct future research. We identified a range synonyms or close semantic relations by a traversal of WordNet[34] (Cognitive Science Laboratory of Princeton University, 2010). The keyword "risk*" did not emerge by this

[34] WordNet is a widely used semantic lexicon designed to support automatic text analysis. The traversal started from the synsets containing the word "uncertain*", and stopped along a given path when we reached a word that we judged inappropriate based on the corresponding synsets. We then included the words encountered at intermediate steps, along with direct negations of their antonyms (e.g. "not clear" cf. "unclear").

method[35] but is surely of interest in our specific domain. Indeed 24% of Factiva FT records containing "uncertain*" also contain "risk*". We therefore included it in our investigation.

The frequency of articles containing the selected keyphrases, as a ratio to the frequency of articles containing "uncertain*", is reported in Table 1. "Risk*" occurred the most frequently, in three times as many articles as "uncertain*", and adding "risk*" to the keyphrase list along with "uncertain*" would more than double m. By contrast, the other keyphrases are much *less* frequent than "uncertain*", with the next most common "unclear" occurring in only around one quarter as many articles. Conversely, articles containing "uncertain*" accounted for more than half of articles containing any of these additional keyphrases (excluding "risk*"). The time series for variants on U, calculated using these different keyphrases on a standalone basis (instead of "uncertain*"), are strongly correlated with our baseline U measure as calculated using "uncertain*". This supports our prior that these keyphrases are strongly semantically related.

Table 8: Additional keyphrases – basic relationship to "uncertain*", 1984–2012

Keyphrase	Ratio of m using only this keyphrase to m using only "uncertain*"		Correlation of a variant of U, calculated using only this keyphrase, to baseline U, calculated using only "uncertain*"			
	standalone	incremental	weekly	monthly	quarterly	yearly
unclear*	0.26	0.23	0.42	0.51	0.56	0.62
unsure*	0.05	0.04	0.24	0.45	0.63	0.81
unpredictabl*	0.08	0.07	0.35	0.59	0.71	0.85
not certain*	0.15	0.01	0.29	0.38	0.61	0.77
not clear*	n/d†	0.14	n/d†	n/d†	n/d†	n/d†
not sure*	0.08	0.07	0.43	0.47	0.66	0.72
not predictabl*	0.00	0.00	0.02	0.05	0.09	0.16
any the above	1.55	0.54	n/d	n/d	n/d	n/d
risk*	2.96	2.38	n/d	n/d	0.65	0.71

Notes: the standalone ratio in the second column is the ratio of the number of articles containing the given keyphrase to the number containing "uncertain*". Articles containing both keyphrases are counted in both numerator and denominator of the ratio. The incremental ratio excludes articles that contain "uncertain*" from the numerator. It represents the proportional increase in the number of articles that would be classified as uncertain, if adding this one keyphrase to our baseline keyphrase set {"uncertain*"}. Correlations are calculated in relation to U rather than to m, to avoid spurious correlation in case article counts are non-stationary. n/d = not calculable from our dataset. † Some figures were not available for "not clear*" due to a data collection error (omitting the wildcard in the database search) that was not easily rectifiable.

In conclusion, future investigations should prioritise the keyphrase "risk*". The incremental effect from adding other synonyms to a keyphrase list that already contains "uncertain*" is likely to be of second-order.

8.2 Disaggregating uncertainty by subject

Disaggregating the news flow and uncertainty references by subject, would allow construction of uncertainty indices customised to particular decision contexts, which could be used to model the effect of uncertainty in those contexts.

Classification of articles by subject can be based either on the subject tags often available in article metadata, or on occurrence of selected keyphrases within the article text. In the latter approach, the

[35] Five of the six senses of "risk" recorded in WordNet, which is designed for general usage and not specifically for economics, refer to possibility or likelihood of loss or negative outcome, rather than either uncertainty *per se* or stochastic risk.

keyphrase list could be chosen either *a priori* or by assembling an index of keyphrases occurring in proximity to uncertainty keyphrases, and filtering for subjects of interest.

In the next Chapter, we use company tags in the article metadata to construct a company-level uncertainty measure, where the subject is anything about the given company. We then use this to model the effect of uncertainty on company capital investment.

References

Alexopoulos, M., & Cohen, J. (2009). *Uncertain times, uncertain measures* (No. 352). *University of Toronto, Department of Economics Working Papers*.

Baker, S. R., Bloom, N., & Davis, S. J. (2013). *Measuring Economic Policy Uncertainty*. Retrieved from www.policyuncertainty.com

Barndorff-Nielsen, O. E., Kinnebrock, S., & Shephard, N. (2008). *Measuring downside risk — realised semivariance*.

Bekaert, G., Engstrom, E., & Xing, Y. (2009). Risk, uncertainty, and asset prices. *Journal of Financial Economics*, *91*(1), 59–82. doi:10.1016/j.jfineco.2008.01.005

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685.

Bloom, N. (2014). Fluctuations in Uncertainty. *Journal of Economic Perspectives*, *28*(2), 153–176. doi:10.1257/jep.28.2.153

Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *The Journal of Finance*, *56*(1), 1–43. Retrieved from http://links.jstor.org/sici?sici=0022-1082(200102)56:1<1:HISBMV>2.0.CO;2-7

Cognitive Science Laboratory of Princeton University. (2010). About WordNet. Retrieved from http://wordnet.princeton.edu

Dendy, C., Mumtaz, H., & Silver, L. (2013). *An uncertainty index for the UK 1986-2012*.

Edwardes, M. D. (1995). A confidence interval for $Pr(X < Y) - Pr(X > Y)$ estimated from simple cluster samples. *Biometrics*, *51*(2), 571–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7662846

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, *82*(397), 171–185.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). The Econometric Society. *Econometrica*, *64*(4), 813–836.

Feunou, B., Jahan-Par, M. R., & Tedongap, R. (2010). *Modeling Market Downside Volatility*.

Granger, C. W. J. (1998). Extracting information from mega-panels and high-frequency data. *Statistica Neerlandica*, *52*(3), 258–272. doi:10.1111/1467-9574.00084

Haddow, A., Hare, C., Hooley, J., & Shakir, T. (2013). *Macroeconomic uncertainty: what is it, how can we measure it and why does it matter? Bank of England Quarterly Bulletin*.

Heber, G., Lunde, A., Shephard, N., & Sheppard, K. (2009). Oxford-Man Institute's realized library, version 0.2. Oxford-Man Institute's realized library. Retrieved January 20, 2015, from http://realized.oxford-man.ox.ac.uk/data

Juselius, K. (2006). *The Cointegrated VAR model: methodology and applications*. *Advanced Texts in Econometrics*. Oxford: Oxford University Press.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Co. Retrieved from http://www.econlib.org/library/Knight/knRUP.html

Kothari, S. P., Li, X., & Short, J. E. (2009). The Effect of Disclosures by Management, Analysts, and Business Press on Cost of Capital, Return Volatility, and Analyst Forecasts: A Study Using Content Analysis. *The Accounting Review*, *84*(5), 1639–1670.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, *17*(3), 1217–1241.

Leahy, J. V, & Whited, T. M. (1996). The Effect of Uncertainty on Investment: Some Stylized Facts. *Journal of Money, Credit and Banking*, *28*(1), 64–83.

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, *LXVI*(1), 35–65.

Mead, N., & Blight, G. (2014). Eurozone crisis timeline. *The Guardian*. Retrieved January 23, 2015, from http://www.theguardian.com/business/interactive/2012/oct/17/eurozone-crisis-interactive-timeline-three-years

Merton, R. C. (1980). On estimating the expected return on the market. *Journal of Financial Economics*, *8*(4), 323–361.

Newson, R. (2002). Parameters behind “nonparametric” statistics: Kendall’s tau, Somers' D and median differences. *The Stata Journal*, *2*(1), 45–64.

Newson, R. (2005). Confidence intervals for rank statistics: Somers’ D and extensions. *The Stata Journal*, *6*, 497–520.

Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data-dependent models for the selection of the truncation lag. *Journal of the American Statistical Association*, *90*(429), 268–281.

Ng, S., & Perron, P. (2001). Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica*, *69*(6), 1519–1554. Retrieved from http://www.jstor.org/stable/2692266

Rossi, B., & Sekhposyan, T. (2015). Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions. *American Economic Review: Papers & Proceedings*.

Schwert, G. W. (1989a). Business cycles, financial crises, and stock volatility. *Carnegie-Rochester Conference Series on Public Policy*, *31*, 83–126.

Schwert, G. W. (1989b). Why Does Stock Market Volatility Change Over Time? *The Journal of Finance*, *XLIV*(5), 1115–1153.

Segal, G., Shaliastovich, I., & Yaron, A. (2014). Good and Bad Uncertainty: Macroeconomic and Financial Market Implications. *Working Paper*, (January).

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *LXII*(3), 1139–1168.

Appendices

A. News-media data

A.1. Readership of the Financial Times

While it is now a global newspaper, with more copies sold abroad than in the UK since September 1998, the FT still has a wider coverage and more detailed analysis of news on UK-listed companies than other national daily newspaper, and it was the unrivalled UK business daily throughout our sample period. It continues to have a large UK readership: the UK print edition had an estimated daily average audience of 319,000, and the FT claimed to have 457,938 unique UK readers across all platforms (including web, mobile etc.) in 2012[36].

Evidence from various surveys suggests that among business decision-makers the FT daily reaches:

- 20% of 1.8 million business 'Purchase Decision Makers'[37] in the UK, and 31% among the sub-segment who 'worked on international business strategies in the past 12 months', which seems likely to correlate with being a key influencer in major investment decisions
- 24% of 435,000 senior business decision leaders across Europe[38]
- 28% of 3,900 senior finance staff of large organisations across the world (both non-financial and banks) responsible for raising finance from capital markets[39].

Among institutional investors the FT daily reaches:

- 36% of senior decision-makers in buy-side financial institutions globally[40]
- 79% of institutional buy-side investors with more than USD100million under management and based in UK/ROI[41].

Finally, compared to other publications, the FT was considered the most credible media owner in the reporting of financial and economic issues, by those who personally managed assets worth USD5billion or more (a universe of 2,522 individuals).

A.2. Canonical set of FT publication days

We constructed a canonical set of FT publication days by cross-referencing daily article counts from Factiva, Nexis UK; monthly article counts from Proquest; selected daily facsimile copies of the print newspaper from Gale; and miscellaneous other sources used to establish reasons for non-publication on certain days (e.g. print stoppages).

The FT was usually published daily on Monday through Saturday throughout the sample period. We counted non-Sunday days with 50 or more articles after de-duplication as publication days. Typically

---

[36] Source: http://www.fttoolkit.co.uk/admediakit/pdfs/adga/Adga_Certificate_April_12_to_March_13.pdf, retrieved on 6 Nov 2013.

[37] Source: British Business Survey (BBS) 2011.

[38] Defined as "C-suites, Head of Department, other senior management and directors/VPs […] who sit in industrial and commercial companies with 250 or more employees, and if company turnover is greater than £40m the threshold is reduced to 150+ employees". Source: Business Elite Europe (BE:EUROPE) 2013 survey. We did not have access to segmentation for the UK

[39] Source: Global Capital Markets Survey (GCMS) 2011.

[40] Source: Global Capital Markets Survey (GCMS) 2011.

[41] Source: Worldwide Professional Investment Community (PIC) Study 2010, Erdos & Morgan Inc.

an FT issue contains over 100 articles, but occasionally slimmer issues are published (e.g. on 27 December in some years). However, on a random subsample of the 76 days with between zero and 49 records in Factiva, all records were misclassified by date, appearing on a consecutive day in the Gale facsimile copy. We discarded all articles on such days rather than attempt to manually re-classify them, given the effort that would be required and the tiny scale of the resulting error in comparison to other potential sources of noise.

Most canonical non-publication days were either Sundays or Bank Holidays, but also included an FT coverage gap in all available electronic databases from 2 June to 8 August 1983 inclusive, and a handful of disparate days with no articles in any of the available databases, for assorted reasons[42].

A.3. Canonical daily total article counts

Due the greater volatility of duplication rates in Factiva, and the impossibility of reliably de-duplicating this total without access to the full text for every FT article (which we did not have, unlike with the subset of articles containing uncertainty keyphrases or tagged with particular companies) we fell back on the Nexis UK daily FT record count as our canonical daily total FT article count, subject to the following adjustments:

- For January 1990 to January 1992, June 1992 to December 1992, and October to November 2007 inclusive, we reduce the Nexis UK count by 20% as an approximate adjustment for the residual duplication observed during these periods.
- On seven publication days when Nexis UK has zero records, we substitute the Factiva record count (always greater than 50 in these cases) adjusted by a multiplying factor equal the linearly interpolated mean of the ratio of Nexis to Factiva records. We do likewise for 10 publication days when Nexis UK has more than zero but fewer than 50 records (Factiva has more than 50 records in all these cases).
- On five days when the canonical count is still between 50 and 70 records, we interpolate in the same way but using item counts and ratios from the Gale and Nexis databases rather than Factiva and Nexis. This is to avoid low (and erroneous) outliers.
- On Saturday 2 January 1999, which is a publication day but has no records in Nexis UK, due to a database error, we use the Factiva record count (which appears to be stable and reliable around that time).

[42] 11 August 1982, 22 September 1982, 27 June 1984 (printers' stoppages). Friday 31 December 1999 (perhaps due to 'millennium bug' concerns and preparations); Saturday 11 April 1998 (Easter); Saturday 9 October 1982, Saturday 26 November 1983, Friday 6 July 1984 (no apparent reason).

B. Unit root tests

Table 9: Unit-root tests on news-media uncertainty, U, 1984m1–2012m12

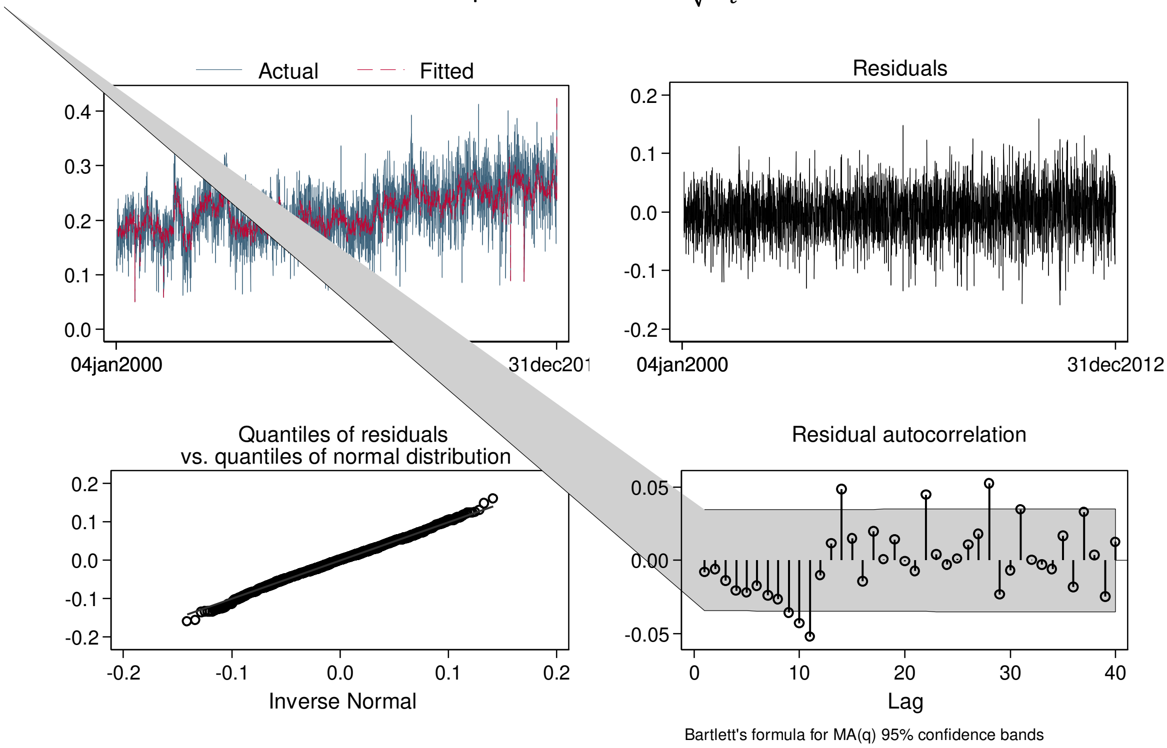
Deterministic trend?	Lag criterion	Lags	α		1%	5%	10%
Linear	Ng-Perron	34	-2.893	**	-3.480	-2.835	-2.547
	Min SC	16	-4.241	***	-3.480	-2.838	-2.550
	Min MAIC	34	-2.893	**	-3.480	-2.835	-2.547
None	Ng-Perron	34	-2.368	**	-2.580	-1.947	-1.623
	Min SC	16	-3.402	***	-2.580	-1.949	-1.625
	Min MAIC	34	-2.368	**	-2.580	-1.947	-1.623

Notes: α is the parameter on the lagged dependent variable in the DF-GLS test of Elliott et al. (1996). The null is that U has a unit root, and the alternative is stationarity or linear trend-stationarity (corresponding to "None" or "Linear" in the "Determinstic trend" column). Lag selection is according to one of three alternative standard criterion: sequential-t algorithm of Ng & Perron (1995), minimum Schwarz information criterion (min SC), and minimum modified information criterion of Ng & Perron (2001). Critical values shown in the last three columns are interpolated as per Stata command −dfgls−. All tests use 7290 observations. *, ** and *** indicate rejection of the null at 10%, 5% and 1% levels respectively.

C. VAR results

Figure 21: Residual diagnostics for daily VAR, 2000–2012

Dependent variable: $\sqrt{U_t}$



Dependent variable: $\ln(\sigma_t)$

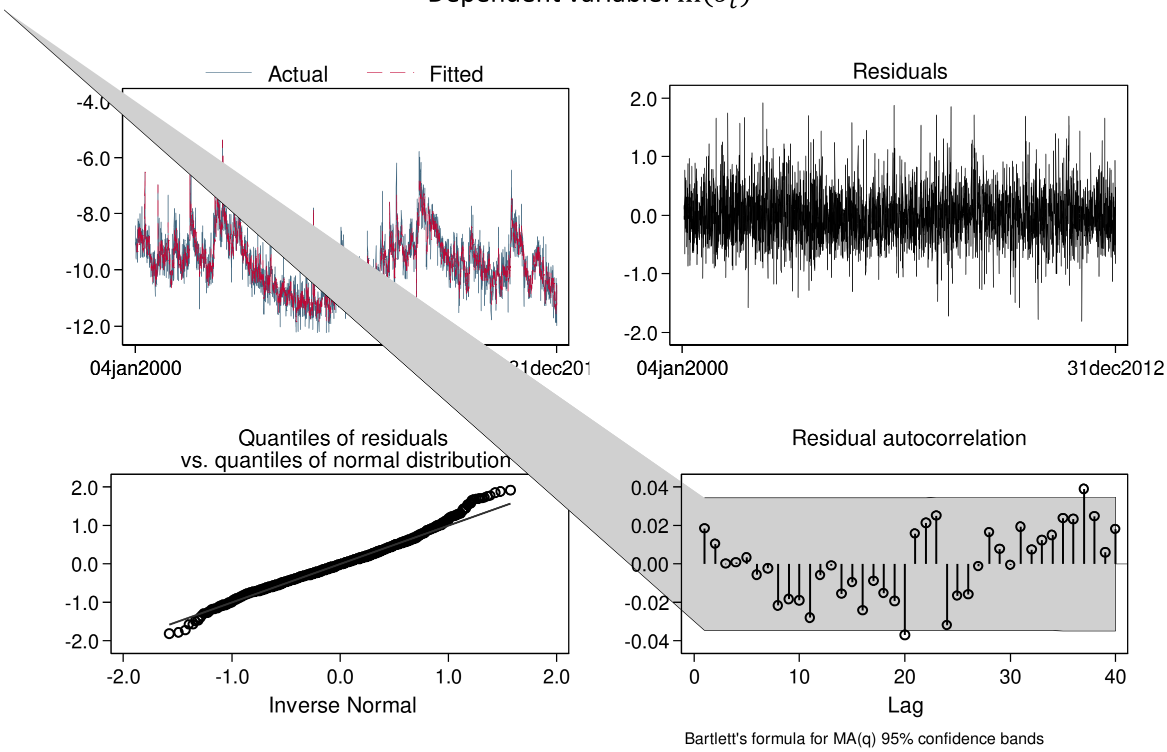
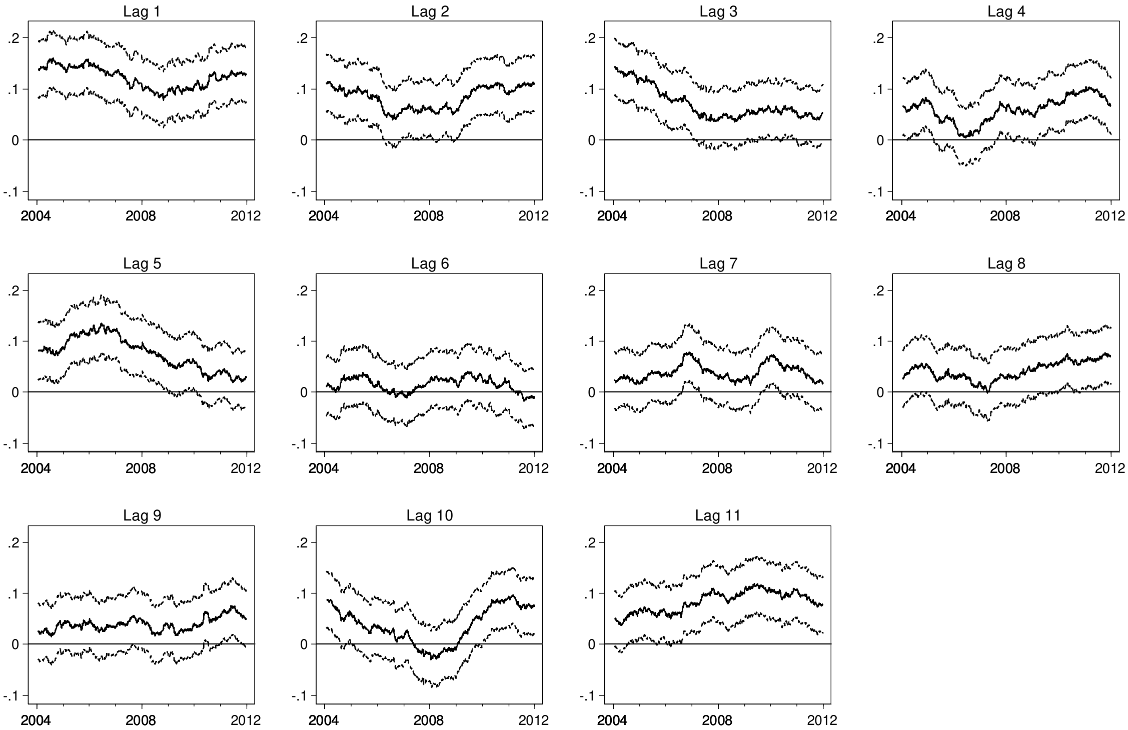


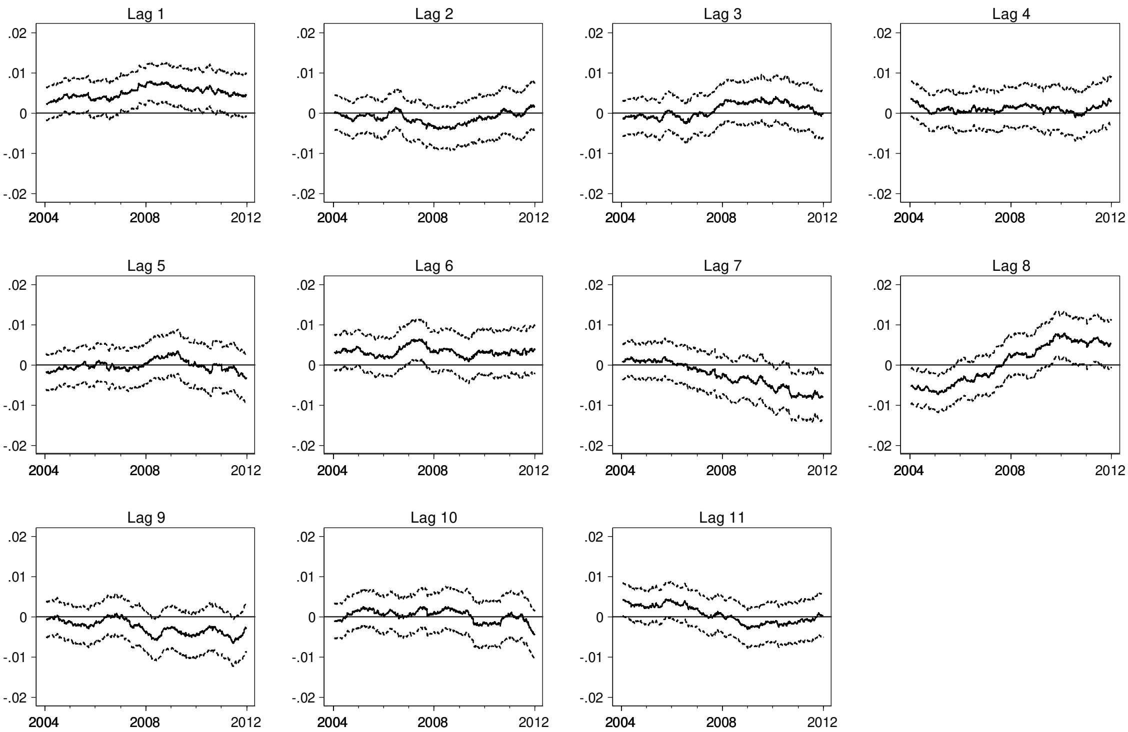
Figure 22: Rolling parameter estimates, daily VAR model, 2000–2012

Dependent variable: *U*

Explanatory variables: lags of U



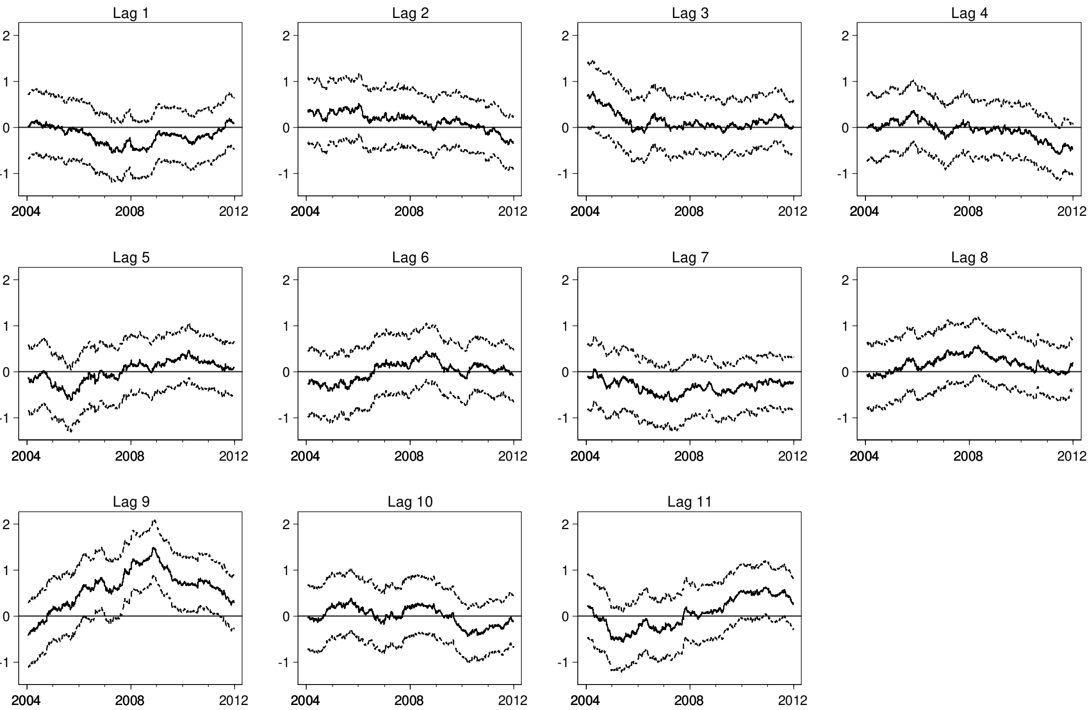
Explanatory variables: lags of σ



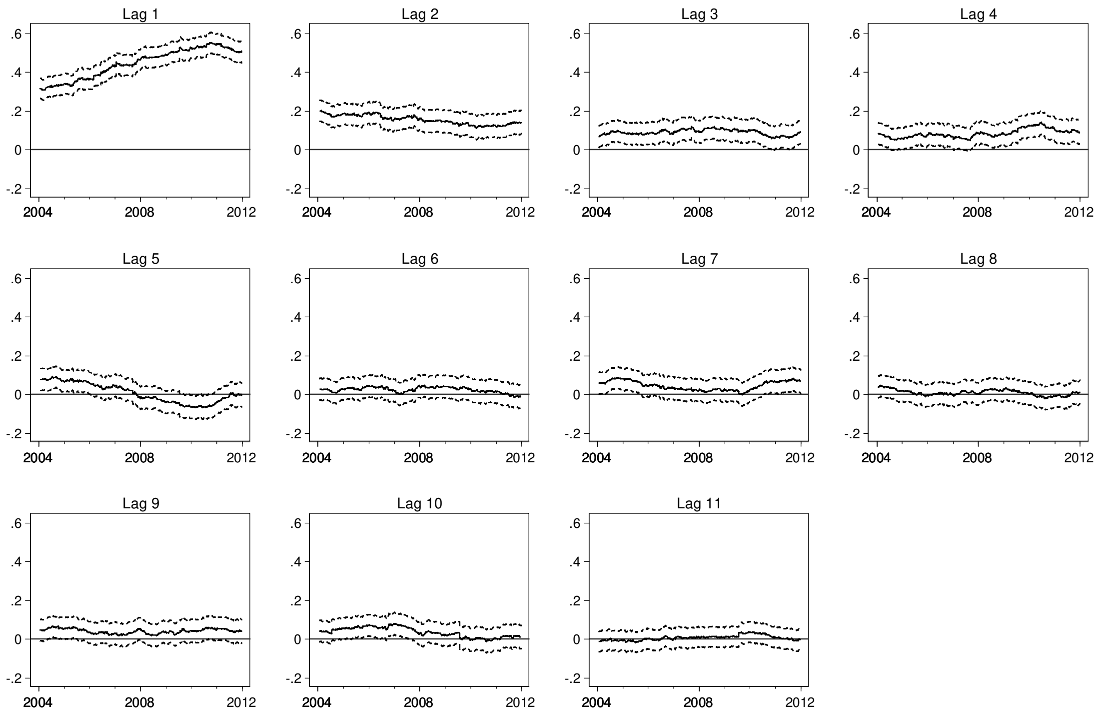
Note: horizontal axis indicates end of five year rolling window.

Dependent variable: σ

Explanatory variables: lags of U



Explanatory variables: lags ofσ



Note: horizontal axis indicates end of five year rolling window.